

# THIS WEEK

## EDITORIALS

**AUSTRALIA** Carbon tax reignites climate politics down under **p.140**

**WORLD VIEW** Criterion from the black lagoon, the sequel **p.141**



**DIGITAL FIRST** How the mole got its extra thumb **p.142**

## There will be blood

*Screening of newborns for genetic disorders is important, but so is educating parents to ensure that they give the proper consent.*

Whether the new arrival is a boy or a girl, most parents do not notice their baby's blood being sampled because it happens so quickly. Within a few days of birth, a health-care worker deftly pricks the heel of the writhing bundle of joy and dabs a few blood spots onto filter paper. The sample is later used to check the newborn for rare genetic disorders.

This is newborn screening in the United States, which identifies problematic conditions in some 3,400 infants each year. In many ways, it is public health care at its best. In others, it demonstrates the sometimes fraught relationship between the medical and research communities and the public.

The problem is not only that many parents miss the heel-prick test at the time, but also that they can be unaware of it altogether. That might be acceptable if the blood sample taken from their child was destroyed after screening, but it is not. In fact, some blood is retained (sometimes indefinitely), stripped of identifying information and used for quality assurance and unrelated biomedical research. Usually this is done without the parents giving their explicit informed consent.

The United States is not the only nation that fails to insist on seeking the written permission of parents before sampling and screening babies. Only a few countries do. But as the News Feature on page 156 shows, the fact that the practice is widespread offers little defence against those who see the lack of consent as a serious problem.

One of those critics is the vociferous campaigner Twila Brase, a former nurse who hates the idea that the government could have the DNA of babies on file. Beneath her inflammatory suggestions that researchers are planning to test infants for a tendency to violence, she has a serious point. Parents must be better informed about the storage and use of these samples.

People want control over their genetic information and that of their children, and they are not getting it. Some parents might have been given a leaflet and some may even have signed a form, but as a recent review shows, state regulation of newborn screening is a mess (M. H. Lewis *et al. Pediatrics* 127, 703–712; 2011).

Only a handful of US states have laws that expressly allow parents to opt out of having blood spots retained for study. At least two changed their regulations only after litigation. And the chance to opt out can be presented at an inopportune time, when parents could struggle to make a clear decision. Although many health officials worry that giving parents a choice will encourage them to refuse the screening, there is evidence that most parents asked for permission will allow their children's blood to be stored for future research (B. A. Tarini *et al. Public Health Genomics* 13, 125–130; 2010). And the potential for biomedical research at a population level is vast. That is why the US National Institutes of Health recently set up the Newborn Screening Translational Research Network, to share data between state repositories and researchers.

To be unclear about how newborn blood is collected and used is the fastest route to turn the public against sampling of newborns for any purpose — including screening programmes. Witness the debacle in Texas, where five million samples had to be destroyed after newspapers published evidence that the state was trying to conceal it was passing on newborn bloodspots to the military.

***“To be unclear about how newborn blood is collected and used is the fastest route to turn the public against sampling.”***

The best solution may be to explain more clearly to parents that the use of baby blood for screening and for research programmes are distinct, and to give them more choice about what is done with samples after screening has occurred. The state of Michigan does this already through its BioTrust for Health programme, which allows

parents to change their minds later.

But there must also be a greater effort to inform and educate parents about the benefits of both newborn screening and biomedical research on anonymous blood spots. To be effective, this cannot be relegated to the period immediately before and after a child's delivery, and should instead be started earlier in pregnancy.

Making these changes will not be easy. The medical infrastructure is complex enough without adding steps and opportunities for people to decide on procedures à la carte. But if parents are not given choice, they will begin to demand it. ■

## REACH further

*Europe's plan for a comprehensive chemical register needs more effort from all involved.*

There are good reasons why European leaders supported moves to tighten the regulation of chemicals by approving the REACH (registration, evaluation, authorization and restriction of chemicals) legislation, which became law in 2006.

The lack of information on how even commonly used substances might harm people and the environment is an internationally recognized problem. REACH is Europe's bold attempt to comprehensively fill this knowledge gap and regulate substances accordingly.

Under the first phase of the legislation, companies from around Europe had to file comprehensive safety data on more than 3,000 substances by December last year. But as we reveal in our News story on page 150, the first independent analysis of the filed data shows that REACH is unlikely to work as planned.

Central to the problems identified by Costanza Rovida, a consultant chemist based in Varese, Italy, is that chemical companies have failed to fill gaps in safety data as required. What's more, European regulators seem to have little leverage to force them to do better.

REACH is often touted as Europe's most complex piece of legislation, but at the moment it looks toothless. Rovida's analysis raises an urgent question: is burdensome and ineffective regulation better than no regulation at all?

There are some parallels with, and perhaps lessons to be drawn from, the difficult birth of Europe's carbon-emissions trading scheme, which, like REACH, has lofty goals — in this case to reduce Europe's carbon footprint. Flaws in the design of the scheme resulted in farce, ridicule and zero reduction in carbon emissions in its first phase, from 2005 to 2007. But subsequent reform made for a stronger second phase, which is now drawing to a close. Details of the third phase are currently being hammered out.

Like REACH, the success of the emissions-trading scheme largely depends on self-reporting by companies — in this case of their annual carbon emissions, for which they get carbon allowances of a corresponding size. Honest participation is encouraged by independent third parties and national authorities that check each submission for accuracy. Hefty fines lurk for those who stray too far from the truth.

It is still early days for REACH, but there are few signs that Europe has the appetite to improve on its dismal first phase in the way it did with emissions trading, or that it will do more than encourage chemical companies to play by the new rules.

Jukka Malm, director of regulatory affairs at the European Chemicals Agency (ECHA), the body responsible for REACH, says he hopes that industry will up its game, but he can do little more than finger wagging to spur it on.

The companies know, as does the ECHA itself, that the agency's biggest weakness is that it has resources to check only a fraction of all data submissions for accuracy and compliance. To deny the regulator the muscle it needs to properly police submissions is a major flaw of the present design of REACH. In addition, there is no immediate threat of legal and financial sanctions against companies that fail to comply.

Nevertheless, REACH has not been a complete flop. It has motivated companies to dig around in their archives and make public old safety data that they had lying around on the substances they produce. It has also got companies to work together and share data.

But these gains are modest in comparison with the new laws' failures. Notably, they have resulted in only small improvements in filling gaping holes in the available data about potentially serious effects of substances on reproduction and development.

And it is also disappointing that the regulations have failed to encourage companies to explore alternatives to animal testing in response to the demand for more data.

What now? If European leaders allow business as usual, they will find themselves under pressure to scrap REACH completely. It will quickly reach a point where its costs, in terms of time and money, heavily outweigh the meagre gain in information.

That would be a missed opportunity. The legislation rightly aims to correct some serious problems, even though Rovida's analysis has identified real difficulties. Europe has until the 2013 deadline for the next phase of submissions to admit that it has a problem, and to reach a little further to solve it. ■

***"Is burdensome and ineffective regulation better than no regulation at all?"***

## Sunday best

*Good news — Australia's politicians have rediscovered climate change.*

Some called it Carbon Sunday; others called it world-leading; yet others called it something much ruder. Either way, last weekend Australia took a bold step forward, announcing a package of measures to tackle the country's disproportionately high greenhouse-gas emissions.

At the heart of the package is a levy on carbon, which will be applied to emissions from the nation's 500 largest polluters. The package, of course, does not come close to cutting emissions by the amount required to head off the worst of global warming — but then, do any concrete political measures announced so far do that?

By such dismal reckoning, it would be easy to dismiss the Australian effort as weak — a mere drop in the ocean — and critics are already doing so. But the package deserves a more sympathetic and considered response. The policy breaks new ground, moves in the correct direction and comes at a welcome time, given how climate change has plummeted down the international political agenda over the past year or so.

The plan, announced by Prime Minister Julia Gillard on 10 July, will see the country's biggest emitters pay a levy of Aus\$23 (US\$25) on each tonne of carbon dioxide that they send into the atmosphere. The price will increase above inflation for 3–5 years, after which the strategy will grow into a broader emissions-trading scheme. Gillard said that by 2020, the move would reduce emissions to 5% below levels seen in 2000 — an overall saving of some 160 million tonnes of carbon. Alongside this introduction of a carbon tax, Gillard increased Australia's long-term emissions target from a 60% cut below 2000

levels by 2050 to an 80% cut. And she pledged Aus\$10 billion to develop renewable energy.

Less than a year ago, climate change was too hot for Australian politicians. Despite opinion polls that showed strong public support for policies to address greenhouse-gas emissions, both the Liberal and Labor parties installed leaders who promised to do less — in fact, before the election last August, Gillard, leader of the Australian Labor Party, had specifically promised not to introduce a carbon tax. But when the votes pushed Labor into a coalition with the Australian Greens and independent members of parliament, the party had to promise its partners that it would embrace a fresh approach to climate change. That compromise produced the weekend's announcement. If policies to restrict emissions can undergo such a resurrection in Australia, they can in other places, too.

Australia's example also gives encouraging signs that climate policies need no longer be proposed by government environment departments, only to be fought against by treasury colleagues who control finances. The Australian action on carbon emissions comes alongside broader reforms of the country's tax system, partly to help ease the burden on citizens affected by the hikes in energy price that are anticipated as a result of the carbon levy. Some of the proceeds from the levy will also be intelligently rechannelled back to the industries affected, to help them adjust to and invest in clean energy. That is the way for a government to hush complaints that action on global warming is being used to raise general tax revenue.

It is far from clear that Gillard's grand plan will be realized. Tony Abbott, leader of the opposition and a professed climate sceptic, has already promised to scrap the scheme, should he win power in the general election set to take place in 2013. He has vowed to make the vote a referendum on the new tax, so we might not have seen the final Carbon Sunday. Still, welcome back, Australia. ■

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunq](http://go.nature.com/xhunq)





## The dubious benefits of broader impact

*Assessments of the wider value of research are unpopular. Proposed changes will only produce more hype and hypocrisy, says Daniel Sarewitz.*

Since 1997, it has not been sufficient for US researchers seeking grants from the National Science Foundation (NSF) to merely explain the intellectual merit of their proposal. They must also justify their work in terms of a variety of ‘broader impacts’.

Politicians worldwide no longer accept that public investments in science automatically bring social benefits. They increasingly expect research expenditure to be justified on its potential contribution to social and economic goals. In the United States, this expectation has resulted in the NSF’s notorious Criterion 2.

Criterion 2 is used by peer reviewers to check that projects will promote education and training, broaden participation, improve infrastructure for research and education, disseminate knowledge or deliver more general social benefits. Yet, according to a review by the National Science Board (the NSF’s advisory and oversight body), the criterion “can be very confusing to the research community, which continues to express frustration in interpreting and thus responding effectively”.

Last month, the board published a revised criterion, and scientists had until this week to provide comments to the NSF before the final version is issued. But Criterion 2.1, as it might be called, is just as confusing and counterproductive as its predecessor.

At the heart of the new approach is “a broad set of important national goals”. Some address education, training and diversity; others highlight institutional factors (“partnerships between academia and industry”); yet others focus on the particular goals of “economic competitiveness” and “national security”. The new Criterion 2 would require that all proposals provide “a compelling description of how the project or the [principal investigator] will advance” one or more of the goals.

The nine goals seem at best arbitrary, and at worst an exercise in political triangulation. How else to explain the absence of such important aims as better energy technology, more effective environmental management, reinvigorated manufacturing, reduced vulnerability to natural and technological hazards, reversal of urban-infrastructure decay or improved performance of the research system? These are the sorts of goal that continue to justify public investments in fundamental research.

Yet, more troubling than the goals themselves is the problem of democratic legitimacy. In applying Criterion 2, peer-review panels will often need to choose between projects of equal intellectual merit that serve different national goals. Who gave such panels the authority to decide, for example, whether a claim to advance participation of minorities is more or less important than one to advance national security?

This problem is exacerbated by issues of expertise. To convincingly assess how a particular

research project might contribute to national goals could be more difficult than the proposed project itself. Neither project leaders nor peer-review panels are likely to have sufficient expertise to really understand a single project’s capacity to connect to a persistent challenge such as increasing the nation’s science literacy or economic competitiveness.

Individual projects are the wrong lever to bring NSF research into line with national goals. It is not surprising, however, that the NSF and the science board made this mistake — the agency’s public image is dominated by the idea of the individual scientist, advancing the frontiers of knowledge. As its website explains, the “NSF’s task of identifying and funding work at the frontiers of science and engineering is not a ‘top-down’ process. NSF operates from the ‘bottom up,’ keeping close track of research around the United States and the world, maintaining constant contact with the research community.”

Yet the NSF has engaged in ongoing organizational experiments over the past 40 years, aiming to overcome the limits of single-investigator, peer-reviewed science. From massive Engineering Research Centers and Science and Technology Centers that address complex, interdisciplinary problems, to small Rapid Response Research grants to get funds quickly to researchers working on urgent questions, and programmes that push university academics to engage seriously in education, the NSF is committed to top-down behavioural modification of the scientific community, often driven by the vision of agency leaders and linked to national challenges such as climate change or emerging opportunities such as nanotechnology.

Motivating researchers to reflect on their role in society and their claim to public support is a worthy goal. But to do so in the brutal competition for grant money will yield not serious analysis, but hype, cynicism and hypocrisy. The NSF’s capacity to meet broad national goals is best pursued through strategic design and implementation of its programmes, and best assessed at the programme-performance level. Individual projects and scientists should be held accountable to specific programmatic goals, not vague national ones. For example, if an NSF initiative aims to provide information for decision-makers, proposals should have to provide evidence that there is actually a customer for the results of the proposed work. Criterion 2 needs to be flexible and tailored to the goals of particular NSF programmes. Otherwise, it will remain confusing and frustrating for scientists and politicians alike. ■

INDIVIDUAL  
PROJECTS ARE THE  
**WRONG LEVER**  
TO BRING NSF  
RESEARCH INTO  
LINE WITH  
**NATIONAL  
GOALS.**

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/yol99x](http://go.nature.com/yol99x)

**Daniel Sarewitz** is co-director of the Consortium for Science, Policy and Outcomes at Arizona State University, and is based in Washington DC.  
e-mail: [dsarewitz@gmail.com](mailto:dsarewitz@gmail.com)

# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## IMMUNOLOGY

### Crystal ball for flu vaccines

People's immune response to an influenza vaccine can be predicted after vaccination from gene-expression signatures.

Bali Pulendran at Emory University in Atlanta, Georgia, and his colleagues measured immune responses and gene-expression changes in the white blood cells of 56 volunteers who received the inactivated vaccine against seasonal flu. Expression levels of 42 sets of 3 or 4 genes were used to predict flu-specific antibody response to the inactivated vaccine. For example, levels of the gene *CaMKIV* were inversely correlated with the antibody responses.

Mice lacking the *CaMKIV* protein produced more flu antibodies after vaccination than normal mice, confirming the protein's role in the immune response to vaccines. *Nature Immunol.* doi:10.1038/ni.2067 (2011)

## DEVELOPMENTAL BIOLOGY

### How the mole got its 'thumb'

Almost all land vertebrates have five fingers, but moles flout this rule. On top of their five digits, the creatures have co-opted a wrist bone to evolve a pseudo-thumb that increases hand-surface area for digging.

Marcelo Sánchez-Villagra at the University of Zürich in

Switzerland and his colleagues tracked key molecular markers in embryos of the Iberian mole (*Talpa occidentalis*; pictured) and the North American least shrew (*Cryptotis parva*), a close relative that lacks the long, sickle-shaped bone. They found increased expression of *Msx2*, a gene that promotes digit development, in the area of the developing mole paw in which a wrist bone becomes elongated. The gene product was absent from this region in the shrew.

The pseudo-thumb is not technically a sixth digit, because it comes from a wrist bone, and develops later than the five true digits. *Biol. Lett.* doi:10.1098/rsbl.2011.0494 (2011)



## ASTRONOMY

### Galactic dust from exploding stars

Supernova 1987A is the remnant of an exploded star located in the Large Magellanic Cloud (pictured), a dwarf galaxy some 49 kiloparsecs from Earth. New observations from the Herschel Space Observatory indicate that the explosion probably generated a mass of dust equivalent to 0.4–0.7 times the mass of the Sun.

Mikako Matsuura at University College London and her colleagues say that the vast amount of dust produced by 1987A lends support to the theory that supernovae generated much of the dust seen in distant galaxies.

*Science* doi:10.1126/science.1205983 (2011)

## NEURODEVELOPMENT

### Defects from stunted neurites

Mutations in the gene *FOXP2* lead to speech and language impairments in humans, but its exact role has not been clear. It turns out that the gene regulates other genes involved in the growth and branching of neuronal projections, making it a key player in neurodevelopment.

Simon Fisher at the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands, and his colleagues screened embryonic mouse brain tissue for genes that the *FOXP2* protein binds to and teased out 264 targets. These genes cluster

in networks that control the formation of neurites, which connect neurons to each other. In mice making defective *FOXP2*, neurons showed reduced neurite outgrowth and branching.

*PLoS Genet.* 7, e1002145 (2011)

## CANCER

### Pushing back on drug resistance

Lung tumours may contain a mix of drug-resistant and drug-sensitive cells. Modified drug regimens could exploit this to delay the emergence of resistant tumours.

Certain non-small-cell lung cancers commonly acquire drug resistance, most often through a mutation called



M. VEIGA/PHOTOLIBRARY

EUROPEAN SOUTHERN OBSERVATORY



T790M in the gene *EGFR*. Franziska Michor at the Dana-Farber Cancer Institute in Boston, Massachusetts, William Pao at the Vanderbilt-Ingram Cancer Center in Nashville, Tennessee, and their team cultured resistant and non-resistant cells and found that those with the T790M mutation grew more slowly than drug-sensitive cells. Populations containing resistant cells also regained sensitivity when drug treatment was withdrawn. This, along with data from clinical trials, led the authors to suggest that some resistant tumours still contain drug-sensitive cells that can repopulate the tumour when a drug is taken away.

The authors incorporated these data in a mathematical model to predict tumour behaviour. They estimate that, after drug withdrawal, an *in vitro* population in which 87.5% of cells are resistant would take 35–40 days to shift to just 1% resistant cells. They propose adding a weekly high-dose drug pulse to the daily low-dose regimen to delay resistance.

*Sci. Transl. Med.* 3, 90ra59 (2011)

#### MATERIALS SCIENCE

### Phase-shifters magnified

Some solids naturally fluctuate between two structural forms; now researchers have followed such a transformation directly at atomic resolution in a copper sulphide nanorod. Understanding this process at the atomic scale might lead to the rational design of novel materials that exploit such transformations, such as memory-storage materials.

Haimei Zheng and Paul Alivisatos at the Lawrence Berkeley National Laboratory in Berkeley, California, and their colleagues used high-resolution transmission electron microscopy to watch copper sulphide nanorods oscillate between two solid-phase structures when heated by an electron beam. The transition occurred just above

room temperature, and the material oscillated a number of times before its structure reached a stable configuration. Defects in the material strongly influenced the dynamics of the transformation by partitioning the nanorod into different domains, each with a different oscillation frequency.

*Science* 333, 206–209 (2011)

#### IMMUNOLOGY

### Virus detected upon entry

Retroviruses such as HIV are notorious for their ability to dodge the mammalian immune system, but researchers have pinpointed a mechanism by which retrovirus-resistant mice detect and respond to retroviral infection.

The body's innate immune system detects pathogens using specific receptors, which then trigger the antibody and cellular responses. Tatyana Golovkina at the University of Chicago, Illinois, and her team infected retrovirus-resistant mice with mouse retroviruses. Virus that had been irradiated with ultraviolet light — and had thus been rendered incapable of replicating — was just as able to elicit an antibody response as nonirradiated virus, suggesting that viral entry is enough to trigger the response.

The authors then homed in on the receptor that senses the viral RNA: TLR7. They suggest that virus-sensing occurs in the cell's endosomes — membrane-bound compartments in which TLR7 resides.

*Immunity* doi:10.1016/j.immuni.2011.05.011 (2011)

#### CLIMATE CHANGE

### The cooling effects of haze

A rise in sulphur emissions attributable mainly to new coal-fired power plants in Asia, and China in particular, may have helped to stabilize global temperatures over the past decade. Sulphur aerosols reflect solar radiation back into space.

Robert Kaufmann at Boston

## COMMUNITY CHOICE

The most viewed papers in science

#### PLANT BIOLOGY

### New light shed on leaf growth

**HIGHLY READ**  
on genesdev.  
cshlp.org  
1 July–5 July

The development of new leaves is triggered by light, a finding that contradicts 150 years of conventional thinking. Leaf initiation was thought to be unaffected by environmental cues such as light because

the shoot apical meristem — the region at the top of the plant stem responsible for new growth — is sheltered by older leaves.

Cris Kuhlemeier and his team at the University of Bern studied leaf initiation in tomato plants grown in light or darkness. New leaves did not grow on plants kept in the dark, but leaf initiation resumed when plants were transferred to the light. Chemically blocking photosynthesis did not affect leaf production, indicating that light itself is needed.

The researchers propose that light stimulates the cytokinin and auxin signalling pathways; however, it remains unclear how these two hormone systems interact in this process.

*Genes Dev.* 25, 1439–1450 (2011)

University in Massachusetts and his colleagues conducted a statistical analysis of factors that enhanced or offset climate warming between 1998 and 2008, including atmospheric circulation trends and greenhouse-gas levels.

The results suggest that rising sulphur emissions, along with natural climate variability, explain the hiatus in warming.

*Proc. Natl Acad. Sci. USA*  
doi:10.1073/pnas.1102467108 (2011)

#### METABOLISM

### Bad fat makes good

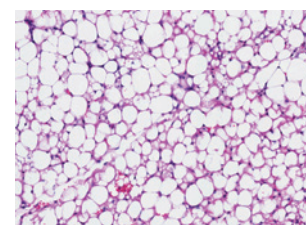
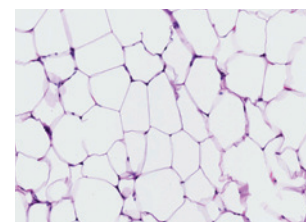
Energy-storing white fat can be converted to energy-burning brown fat by suppressing a cell-signalling pathway.

Sushil Rane at the National Institute of Diabetes and Digestive and Kidney Diseases in Bethesda, Maryland, and his group report that mice lacking the protein SMAD3 are more sensitive to insulin and gain less weight on a high-fat diet than normal mice. The authors found that loss of SMAD3 caused white fat (pictured top) to take on certain key features of brown fat (bottom), such

as the generation of more mitochondria, which power the cell.

SMAD3 is part of the same pathway as the protein TGF $\beta$ . Administering TGF $\beta$  to normal mice blocked the conversion of white to brown fat, whereas inhibiting TGF $\beta$  in mice prone to obesity and type 2 diabetes suppressed both conditions. Furthermore, a survey of 184 non-diabetic humans revealed a correlation between TGF $\beta$  levels and body-mass index.

*Cell Metab.* 14, 67–79 (2011)



ELSEVIER

**NATURE.COM**

For the latest research published by Nature visit:

[www.nature.com/latestresearch](http://www.nature.com/latestresearch)



# SEVEN DAYS

The news in brief

## POLICY

### Price on carbon

Australian Prime Minister Julia Gillard has proposed a carbon dioxide tax of Aus\$23 (US\$25) per tonne for the country's top 500 emitters. The tax would come into effect from 1 July 2012, and the price would increase above inflation until an emissions-trading scheme replaced it in 2015. Gillard now faces a battle to convince Australian voters of her climate-policy package, announced on 10 July, although it already has enough parliamentary support to become law. See page 140 and [go.nature.com/aimkho](http://go.nature.com/aimkho) for more.

### Testing embryos

Germany's parliament has agreed to legalize some genetic testing on embryos generated by *in vitro* fertilization before they are implanted into a prospective mother's uterus. The vote, on 7 July, brings the country closer in line with the United States and most Western European countries, where preimplantation diagnosis of genetic abnormalities has been legal since the 1990s. See [go.nature.com/yoqdpw](http://go.nature.com/yoqdpw) for more.

### James Webb threat

Republicans in the US Congress are threatening to cancel the James Webb Space Telescope (JWST),

## NUMBER CRUNCH

# 164.79

The number of Earth years it takes for the planet Neptune to orbit the Sun. On 11 July, Neptune completed its first orbit since it was discovered on 23 September 1846.



B. INGALLS/NASA

## The shuttle's last launch

"*Atlantis* is flexing its muscles one final time," said NASA's television commentator as the space shuttle climbed away from Earth on 8 July. It is the 135th and last flight of NASA's

shuttle programme. *Atlantis* is currently docked with the International Space Station, where its four crew members are delivering supplies and spare parts, and is due to return on 20 July.

the successor to the Hubble Space Telescope. On 7 July, a House subcommittee voted to cut NASA's funding by US\$1.6 billion (from last year's enacted levels) and to strip the JWST of funding altogether. The telescope's steadily climbing price tag — currently estimated at \$6.5 billion — is devouring NASA's astrophysics budget. The vote was just the first step in a long legislative process, but comes as a warning shot to JWST supporters. See [go.nature.com/c57xiw](http://go.nature.com/c57xiw) for more.

### Nuclear checks

Japan's 54 nuclear power stations are to undergo stress tests in the wake of multiple meltdowns at the Fukushima Daiichi nuclear plant, the government announced last week. The tests will be similar to ones now under

way in the European Union, and will consider the plants' preparedness for earthquakes and floods. As a result of the March earthquake that triggered the meltdown at Fukushima, only 19 of Japan's reactors are operating. It is unclear whether the stress testing will further delay restarts.

### China's oil spills

China's State Oceanic Administration (SOA) has been criticized for its delay in telling the public about two oil spills that have badly polluted 840 square kilometres in the Bohai Gulf, off the country's northeastern coast. On 5 July, the SOA held its first government briefing about the oil leaks, which occurred on 4 and 17 June. Oil spills are supposed to be immediately reported to the public. China's

offshore oil-drilling industry is booming, leading to fears that June's accidents may be followed by many more spills. See [go.nature.com/zf8bjc](http://go.nature.com/zf8bjc) for more.

### Europe's GM ruling

The European Parliament has voted to let European Union member states decide for themselves whether to ban genetically modified crops. Grounds for a ban would include public opinion or socioeconomic and environmental factors (including lack of data on potential environmental harm). The autonomy allowed in the 5 July vote goes further than that proposed by the European Commission, which had suggested that countries should be allowed to decide on all but health and environmental grounds, which

should be left to assessments by the European Food Safety Authority. Member states will discuss the text in October. See [go.nature.com/ecisdw](http://go.nature.com/ecisdw) for more.

## RESEARCH

## Synthetic organs

Surgeons have successfully replaced the windpipe of a cancer patient with a wholly synthetic trachea — showing that donors may not be needed for tracheal transplants. The patient, Andemariam Teklesenbet Beyene — a student of geology at the University of Iceland, Reykjavik — was discharged from the Karolinska University hospital in Stockholm on 9 July, a month after the transplant operation. Stem cells from his bone marrow were grown on a trachea-shaped scaffold of a nanocomposite material developed at University College London. See [go.nature.com/zvuxcd](http://go.nature.com/zvuxcd) for more.

## Telescope damage

A leak of more than 100 litres of coolant has shut down the 8-metre Subaru telescope. The telescope, Japan's largest, is perched atop Hawaii's Mauna Kea mountain. The telescope's interim director, Hideki Takami, says that the instrument will be out of



commission for at least two weeks. Coolant from cables in the telescope's top section leaked onto its main mirror (pictured, marbled with orange coolant) and various electronic instruments, on 2 July. See [go.nature.com/kg2usy](http://go.nature.com/kg2usy) for more.

## Endangered tuna

Some of the world's most commercially valuable fish species should be classed as globally endangered, a study reported on 7 July. Researchers assessed 61 species of scombrids and billfish, applying criteria used by the International Union for the Conservation of Nature (IUCN) in Gland, Switzerland, to produce its Red List of endangered species. Seven species — mostly tuna or marlin — are vulnerable, endangered or critically endangered, they found (B. Collette *et al.* *Science* doi:10.1126/science.1208730;

2011). The assessment was published just before a key meeting on tuna catches in La Jolla, California, on 11–15 July. See [go.nature.com/llhcf7](http://go.nature.com/llhcf7) for more.

## A pinch of salt

A meta-analysis has questioned the oft-repeated connection between the consumption of too much salt and the development of cardiovascular disease. The study, published on 6 July (R. S. Taylor *et al.* *Am. J. Hypertens.* doi:10.1038/ajh.2011.115; 2011), examined the results of seven clinical studies and found no solid proof that reducing salt consumption prevents heart conditions. "Whilst intuitively reducing salt across the board appears to be a good thing, I would say we still need the evidence to prove it," says Rod Taylor, a statistician at the University of Exeter, UK, and the study's leader. See [go.nature.com/mszubs](http://go.nature.com/mszubs) for more.

## PEOPLE

## Journal chief

A nascent open-access life-sciences journal from three major biomedical research funders will be edited by cell biologist Randy Schekman. The Howard Hughes Medical Institute (HHMI) said on 11 July that Schekman, who has been editor of the *Proceedings of the National*

## COMING UP

## 16 JULY

NASA's Dawn spacecraft should reach Vesta, one of the largest bodies of the asteroid belt (see page 147).  
[go.nature.com/gga1b2](http://go.nature.com/gga1b2)

## 17–20 JULY

The International AIDS Society holds its biennial conference on HIV treatment and prevention, in Rome, Italy.  
[go.nature.com/wnr8b5](http://go.nature.com/wnr8b5)

## 21–27 JULY

More results from the Large Hadron Collider and the Tevatron are expected at a conference on high-energy physics in Grenoble, France.  
[go.nature.com/7fyj1b](http://go.nature.com/7fyj1b)

Academy of Sciences since 2006, will begin his new job in August. The as-yet-unnamed online journal, expected to launch next year, is the brainchild of the HHMI, the Max Planck Society and the Wellcome Trust. See [go.nature.com/p1h5vl](http://go.nature.com/p1h5vl) for more.

## FUNDING

## Oil-spill study

The US National Institutes of Health (NIH) will spend US\$25.2 million over the next five years studying the effects of the 2010 Deepwater Horizon oil spill on the health of the general population in the area. The study, announced on 7 July, will focus on vulnerable populations including pregnant women, children, immigrants, fishermen and minorities. It complements an existing NIH project to monitor the health of 55,000 oil-spill clean-up workers and volunteers for up to 10 years.

➔ [NATURE.COM](http://NATURE.COM)

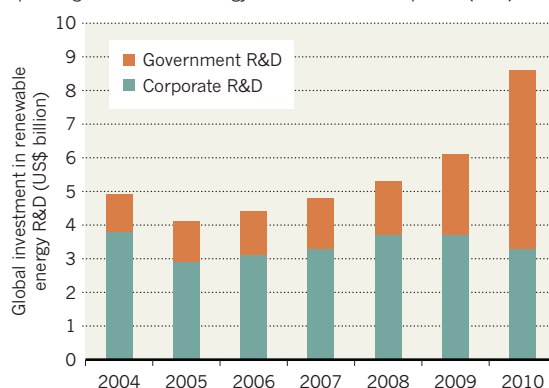
For daily news updates see:  
[www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

Global investments in renewable energy in 2010 jumped by 32% from 2009, to US\$211 billion. The fraction spent on research and development in the sector also rose, by 41% to \$8.6 billion. Although corporate research spending dropped (see chart), government support soared, owing to the 'green stimulus' packages announced in the throes of the financial crisis. Solar power, up by 8% to \$3.6 billion, commanded the largest share of research funding. See [go.nature.com/eftyjh](http://go.nature.com/eftyjh) for more.

## STATE SPENDING BOOSTS RENEWABLES RESEARCH

'Stimulus' packages mean that state spending has overtaken private spending on renewable-energy research and development (R&D).



# NEWS IN FOCUS

**EPIDEMIOLOGY** Mysterious nodding syndrome spreads in Africa **p.148**

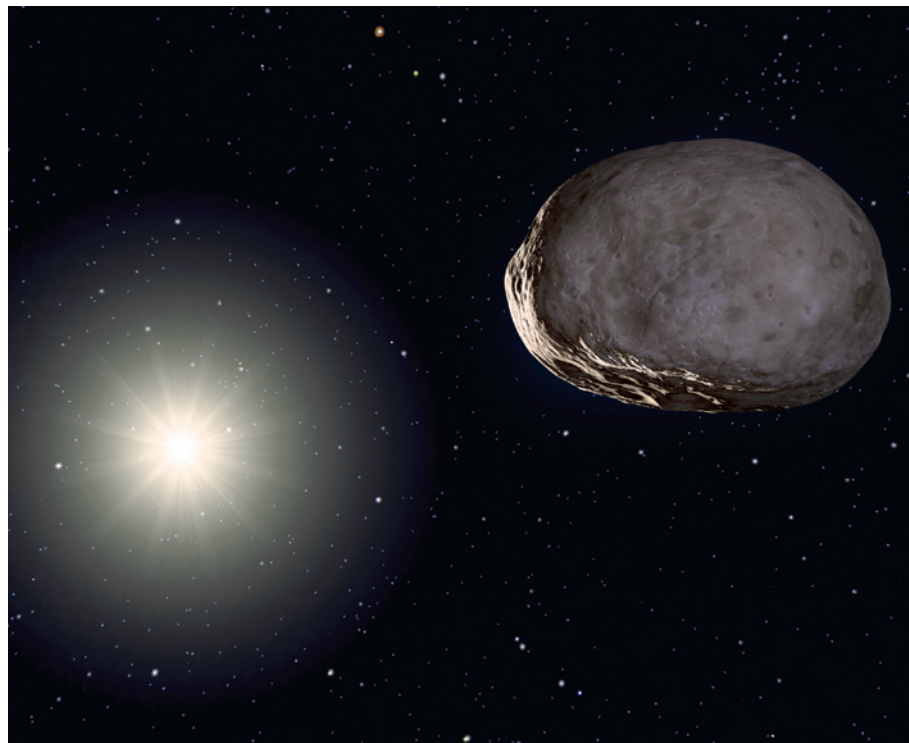
**TOXICOLOGY** Data gaps hobble Europe's ambitious regulations **p.150**

**JAPAN** Study says farming can resume near Fukushima **p.154**



**PUBLIC HEALTH** Attacks on infant screenings draw blood **p.156**

C. BUTLER/SPL



The asteroid Vesta, seen here in an artist's model with the Sun in the background, has complex geology.

## PLANETARY SCIENCE

# Dawn nears Vesta

*Mission poised to explore the Solar System's largest asteroids in detail.*

BY RON COWEN

**T**he Dawn spacecraft had a difficult birth: it was threatened by cost overruns and technical concerns, cancelled, reinstated and scaled down. Now, after a four-year journey spiralling out from Earth's orbit, the probe is set to explore the beginnings of the Solar System.

On 16 July, Dawn will enter orbit around Vesta (see 'Dawn patrol'), an asteroid that, at 500 kilometres wide, is the second largest in the Solar System. It will spend a year there before flying on to Ceres, the Solar System's largest asteroid at nearly 1,000 kilometres wide. There are hundreds of thousands of bodies in the

main asteroid belt, which sprawls between the orbits of Mars and Jupiter and is a storehouse of material that formed early in the Solar System's history. But because Vesta and Ceres have apparently survived in one piece since then, "they are like time capsules telling us about the earliest stages of planet formation", says Carol Raymond, deputy principal investigator of the mission and a planetary scientist at NASA's Jet Propulsion Laboratory in Pasadena, California.

Dawn's comparative study of the two bodies should also help to show how similarly sized objects can evolve very differently. Glimpses of Vesta suggest that its structure is like that of a miniature Earth, with a metallic core and

a rocky mantle and crust, but that its growth was halted when Jupiter's far-reaching gravitational influence prevented asteroids in the belt from coalescing any further. Vesta's composition, deduced from afar through its spectral properties, suggests that after its formation, the asteroid was initially hot enough for lava to ooze out onto its surface. By contrast, Ceres contains many water-bearing minerals, and with an average density lower than that of Earth's rocky crust, it may be one-quarter ice beneath its dust-coated surface. The asteroid could even hold a subsurface ocean, long frozen or perhaps still liquid.

Dawn will use three instruments to probe those differences. A camera will image surface features as small as 10 metres across; a spectrometer will map crustal minerals at various electromagnetic wavelengths; and a  $\gamma$ -ray and neutron detector will reveal the quantities of elements by detecting radiation and particles produced when cosmic rays hit atomic nuclei on the surface of the asteroids.

This information, together with models of where in the early Solar System Ceres and Vesta originated, might confirm one theory as to why the asteroids are so different: that Vesta formed a few million years before Ceres. That would have given Vesta enough time to incorporate the radioactive isotope aluminium-26, which was abundant in the earliest years of the Solar System but decayed before most of the asteroids in the belt formed. The radioactivity could have provided enough heat to drive volcanic eruptions, changing the character of Vesta's surface.

Tectonic upheavals erased evidence of early heating on Earth and the other rocky planets — but not on Vesta. "Vesta is telling us what the planet-formation process looked like after the first 10 minutes in the oven," says Richard Binzel, a planetary scientist at the Massachusetts Institute of Technology in Cambridge and a long-time observer of Vesta, who first tracked the asteroid as part of a school project in 1973.

Dawn will also take advantage of a window into Vesta's interior, notes Christopher Russell, lead scientist of the mission and a geophysicist at the University of California, Los Angeles. Pictures taken by the Hubble Space Telescope in

1996 revealed an impact crater 13 kilometres deep, gouged into the asteroid at its south pole. Dawn will peer into that hole to discern any geological

**NATURE.COM**  
For more on trips to small bodies in the Solar System, visit: [go.nature.com/bbzsho](http://go.nature.com/bbzsho)



► diversity exposed by the impact. Three types of meteorite found on Earth — eucrites, howardites and diogenites — are thought to be chips of Vesta, blasted away by the collision. Linking these convenient specimens to particular internal layers of Vesta is a key driver of the Dawn mission, notes Binzel.

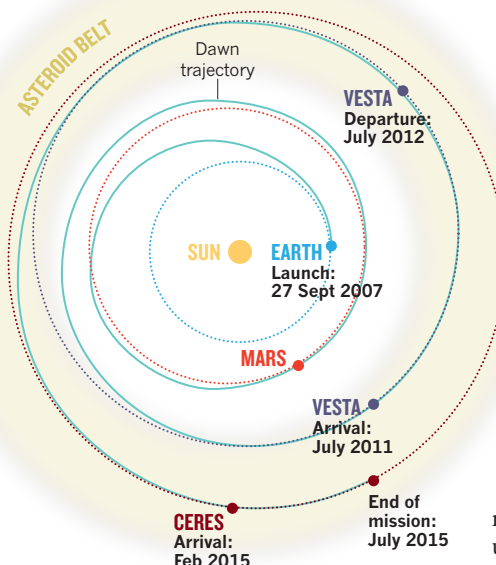
“It’s a little bit like the Humpty Dumpty problem — we’ve got a lot of pieces of Vesta and we’d like to see how they all fit together,” he says.

After its tour of Vesta, Dawn will fire up its ion thrusters — solar-powered jets that supply a weak but long-lasting push — and set a course for Ceres, which it will inspect over five months in 2015.

Before launch, budget issues caused the mission team to drop two instruments originally meant to fly aboard Dawn; one of them,

## DAWN PATROL

A seven-year flight plan includes encounters with two major asteroids.



a magnetometer, will be especially mourned once the craft reaches Ceres. The magnetometer could have looked for fluctuations in the strength of the asteroid's magnetic field that might have provided clues as to whether the body harbours a briny ocean. Losing the instrument “was a big blow,” says Raymond.

Although Dawn has so far survived the ravages of budget changes, politics and four years in interplanetary space, Russell says that he won't relax until the craft enters orbit around Ceres. Casey Lisse, a planetary scientist at Johns Hopkins University's Applied Physics Laboratory in Laurel, Maryland, agrees. “We've learned most of what we can from remote observations of Ceres, and we need an up-close and personal look,” he says. ■

## EPIDEMIOLOGY

# African outbreak stumps experts

*With few leads to go on, researchers pursue the childhood malady nodding syndrome.*

BY MEREDITH WADMAN

The boy was perhaps seven or eight, although he could have been older: among other things, the disease that afflicts him stunts growth. When a seizure began, his mother summoned Sudhir Bunga, who found the boy sitting under a tree in a school playground. “The child was staring blankly and his head was intermittently nodding every five to eight seconds,” Bunga says. “This lasted about three minutes.”

Bunga was not surprised by what he saw. A physician and epidemiologist with the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, he was in rural southern Sudan in May as part of an emergency-response team trying to assess a mysterious illness seen in children in the region. But despite his preparation, Bunga was deeply affected by his first encounter with ‘nodding syndrome’. “Actually seeing it out in the community was overwhelming and distressing,” he says. “The burden of the disease looked really high.”

Nodding syndrome is a poorly understood and seemingly growing problem in eastern Africa, where it is devastating communities in South Sudan and northern Uganda. It has existed separately for decades in a secluded mountainous area of southern Tanzania<sup>1</sup>. In South Sudan, “it’s affecting thousands of children,” says Abdinasir Abubakar, a physician for the World Health Organization (WHO)

based in South Sudan who coordinated the recent CDC trip. “Of course, the question is whether this syndrome is spreading to new communities.”

For South Sudan, which achieved political independence only on 9 July, the syndrome raises the additional fear that the new nation's limited capacity to deal with an emerging medical threat will be quickly overwhelmed without outside resources and expertise.

“Nodding syndrome cannot be left with the nascent government in South Sudan,” says

Martin Opoka, an epidemiologist with the WHO's eastern Mediterranean regional office in Cairo. “They will certainly need assistance from the international community.”

Opoka helped to investigate the occurrence of nodding syndrome in southern Sudan as part of a WHO team in 2002, and returned to the region this year to assist the CDC investigators. The CDC team — consisting of four physician-epidemiologists with specialties in paediatrics, neurology and nutrition — was dispatched by the US agency's Division



In some villages in South Sudan, almost every family has a child affected by nodding syndrome.



Companies must assess the toxicities of their products, but are there enough data in the pipeline?

#### TOXICOLOGY

# Data gaps threaten chemical safety law

*European companies are not providing robust information to regulators or alternatives to animal experiments.*

BY NATASHA GILBERT

Europe's sweeping chemicals law, sometimes described as its most complex piece of legislation, was meant to regulate thousands of common substances to protect people and the environment from harm. But four years after REACH (registration, evaluation, authorization and restriction of chemicals) came into force, the burdensome, costly law is beginning to look strangely toothless. Evidence seen exclusively by *Nature* shows that companies have failed to fill gaps in safety data — and European regulators have done little to pressure them.

REACH requires companies that produce or sell chemicals in the European Union to register toxicity data on the compounds and outline any new tests needed to clarify their biological effects, especially on reproduction and the development of offspring. Before

REACH, these costly tests — multigenerational rat studies can cost up to €2 million (US\$2.8 million) per chemical — were rarely performed in Europe because the previous law required them only for substances produced in very large quantities. Switching to the REACH system was predicted to trigger millions of extra animal tests (see *Nature* 460, 1065; 2009), so companies were also expected to propose alternative methods wherever possible to minimize the use of animals.

The legislation requires companies to compile all safety information and planned tests into dossiers, one for each chemical, and submit them to REACH's regulator, the European Chemical Agency (ECHA), based in Helsinki. The ECHA has little power to enforce the regulations, however, leaving any penalties for non-compliance to individual governments.

Dossiers for more than 3,200 of the most ubiquitous chemicals have been filed with the

agency, with more to come over the next seven years.

Costanza Rovida, a consultant chemist based in Varese, Italy, has now analysed summaries of 200 of these dossiers, chosen at random. She plans to analyse a further 800 summaries and present the findings at the 8th World Congress on Alternatives and Animal Use in the Life Sciences in Montreal, Canada, in August. But already, Rovida has uncovered a host of problems.

Commissioned by the European arm of the Center for Alternatives to Animal Testing (CAAT) at the University of Konstanz, Germany, her research shows that many dossiers rely heavily on old data and fail to suggest new tests, and that few include any mention of non-animal testing methods (see 'Mind the gap'). The ECHA acknowledges that there is room for improvement. "Industry has not taken full responsibility for the quality of data," says Jukka Malm, director of regulatory affairs at the ECHA.

The agency plans to check all dossiers that include proposals for new animal studies — but will look at only 5% of those that have no test proposals, a stipulation set out in the law. To some observers, this hands-off approach highlights a potential weakness in the system. "The purpose of REACH is to get data on many chemicals," says Thomas Hartung, director of the CAAT at its US headquarters in Baltimore, Maryland. "But it is clear that industry wants to avoid testing." If only 5% of dossiers that do not propose tests are checked, "we will not really get a lot of new information," he says.

#### CREATIVE APPROACH

Rovida found that roughly one-third of the dossiers provide animal data on reproductive and developmental toxicity. But much of the information is from old studies — some more than 20 years old — "that don't meet today's testing standards," says Rovida.

Given the existing paucity of animal data on reproductive and developmental toxicity, toxicologists had expected many of the dossiers to propose new studies. However, Rovida says that her analysis shows that few new tests are being proposed to reproduce or challenge the findings. Some 36% of the dossiers she looked at fail to make conclusive judgements about the chemical's reproductive or developmental toxicity (see 'Chloroaniline') — but only 7% and 7.5%, respectively, propose new animal studies to clarify these effects.

Sebastian Hoffmann, a toxicologist based in Cologne, Germany, who works as an industry consultant on REACH, says that companies seem to have been "creative" in interpreting REACH's demands for them to fill data gaps.

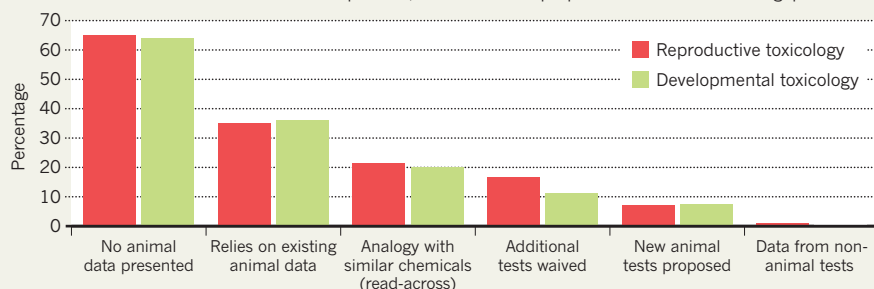
Rovida also found that of the 200 dossiers she examined, only two had provided data from non-animal tests. "This shows that they are not serious about alternative methods," she says.



SOURCE: C. ROVIDA

## MIND THE GAP

About two-thirds of the 200 REACH chemical-safety information dossiers examined do not contain data from animal tests of the chemical in question, but even fewer propose tests to fill that data gap.



"This is not what I hear from our companies," counters Erwin Annys, director of REACH and chemicals policy for the European Chemical Industry Council (CEFIC) in Brussels. "We are not sure that the findings on this small sample are representative," he says. "We remain active in promoting non-alternative testing methods and are looking to get a better understanding by more companies on this issue."

Robert Kavlock, director of the National Center for Computational Toxicology at the US Environmental Protection Agency, says that companies are in a bind because few non-animal testing methods are "scientifically acceptable or ready for regulatory use".

Rovida also found that companies are relying heavily on a technique known as read-across, in which the effects of a substance on human health are predicted by considering the effects of structurally similar chemicals. The REACH legislation, and guidance from the ECHA, is generally supportive of this, as long as it provides sufficiently convincing conclusions.

For around 21% of the dossiers studied by Rovida, reproductive toxicity was judged solely using read-across methods. Although read-across may be appropriate for simple chemical and physical properties, toxicologists are far less positive about its validity for assessing reproductive and developmental toxicity, especially in the absence of other animal test data on the substance. "Whether read-across will prove to be robust is an open question," says Alan Boobis, a toxicologist at Imperial College London. "It will come down to companies proposing good arguments for why read-across is sufficient to make a judgement. But if there are no animal data, I don't know how they can make a case."

The legislation does allow companies to suggest waiving reproductive and developmental toxicity tests, but only if people are unlikely to be significantly exposed to the substance, or if it is already known to damage DNA or gametes. In the dossiers studied by Rovida, companies suggested waiving these tests for 16.5% and 11% of substances, respectively. "Waiving

is quite broadly applied, even though the guidance is extremely strict about when its use is valid," says Hartung.

REACH is far from being useless, emphasizes Rovida. It has forced companies to collate a great deal of existing information about the chemicals they handle, which is an improvement on the situation before REACH. But as a mechanism for collecting and generating data on reproductive and developmental toxicity, it is "a complete failure", she says. What's more, "there is no effort to promote alternative methods. Very little is done to avoid some animal tests," Rovida says.

Hartung hopes that the revelations will build momentum to develop alternative non-animal tests. But Boobis predicts that many more *in vivo* toxicity tests are inevitable. "We have seen this in other areas, where, despite a commitment to reduce animal use where possible, the need to protect public health overrides the lag in scientific development of credible alternatives," he says.

On 30 June, the ECHA published a progress report on REACH that echoes some of Rovida's findings. "The quality of many of the chemical safety assessments is of concern," the report says. In particular, it notes that the quality of the scientific arguments put forward by industry to justify using read-across, and to waive additional safety tests, is "not high enough".

The European Commission, which was involved in drawing up the REACH policy, says that the overall message of the ECHA's report is that the system is working well. "Most of the issues raised in the report can be improved by more efficient implementation," a spokesperson told *Nature*.

## CASE STUDY

## Chloroaniline

The summary dossier on 2-chloroaniline, an aromatic amine used to manufacture pesticides and pharmaceuticals, reports toxicity data that it describes as "conclusive", but says that they are "not sufficient" to classify the substance as toxic to reproduction. It also says that data are "lacking" on whether toxic effects can be passed on to offspring through the mother's milk.

Yet the summary suggests that children of fathers exposed to aromatic amines before conception are at higher risk of brain tumours (J. R. Wilkins and T. Sinks *Am. J. Epidemiol.* **132**, 275–292; 1990). It also cites what Costanza Rovida, a consultant chemist in Varese, Italy, describes as a "robust" (but unpublished) study showing that the substance caused malformations in rats.

The summary dossier does not suggest new tests to investigate 2-chloroaniline. **N.G.**

Although the ECHA lacks enforcement powers, it can ask companies to provide more toxicity data and request new studies if it judges dossiers to be incomplete. If companies don't comply, the agency can report them to national authorities. "Companies are now waiting to see if the ECHA tells them to do extra studies," says Hartung. He argues that the ECHA should check all the dossiers it has received. But Malm says that the ECHA does not have enough resources to check more than 5% of dossiers without test proposals — and even that will be a challenge. Instead, the ECHA will ask industry to "take a serious look back at the quality and improve it proactively rather than wait for us to do compliance checks".

"It is not a failure of REACH," Malm adds. Because this is the first phase of REACH's implementation, and companies are still learning the system, "we should have expected a lower quality of dossier to start with." ■

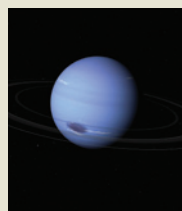
SEE EDITORIAL P.139



## MORE ONLINE

## TOP STORY

Neptune gives up its secrets, including the length of its day.  
[go.nature.com/1kwe81](http://go.nature.com/1kwe81)



## OTHER NEWS

- Sudan splits and science community divides. [go.nature.com/j6md3u](http://go.nature.com/j6md3u)
- Mice with human livers deal with drugs the human way. [go.nature.com/h97kyy](http://go.nature.com/h97kyy)
- Qatar sets sights on stem cells. [go.nature.com/zebknc](http://go.nature.com/zebknc)

W. MYERS/SPL



## ETHICS

# Paxil study under fire

*Trial researcher alleges paper exaggerated antidepressant benefits.*

BY MEREDITH WADMAN

The contentious issue of drug-industry influence over medical-research writing erupted on the campus of the University of Pennsylvania in Philadelphia this week. A professor of psychiatry has alleged that several colleagues — including the chair of his department — allowed their names to be added to a manuscript while ceding control to the global pharmaceutical giant GlaxoSmithKline (GSK). The professor, Jay Amsterdam, also claims that the manuscript, written with an unacknowledged contractor paid by GSK, unduly promotes the company's antidepressant drug Paxil (paroxetine), the subject of the study.

"The published manuscript was biased in its conclusions, made unsubstantiated efficacy claims and downplayed the adverse-event profile of Paxil," Amsterdam's lawyer wrote in an 8 July letter to the Office of Research Integrity (ORI), the body responsible for investigating research misconduct in US Public Health Service agencies and its grant recipients.

The letter accuses the study's academic authors of engaging in scientific misconduct by allowing their names to be attached to the manuscript (C. Nemeroff *et al.* *Am. J. Psychiatry* **158**, 906–912; 2001), which has been cited more than 250 times. Documents accompanying Amsterdam's complaint are offered as evidence that "most if not all" of the authors were handpicked by GSK, working in conjunction with the medical-communications company Scientific Therapeutics Information (STI) in Springfield, New Jersey, to lend credibility to a result that Amsterdam says places Paxil in an overly favourable light. In one such document, Karl Rickels, a psychiatrist not involved with the study who looked at the issue for the department in 2001 said that "apparently ... [academic] participants never had a chance to review or even just see the manuscript before it went to press".

"It has always been GSK's policy and practice for the primary author(s) to have final approval on manuscripts," the company says. "The proper use of medical writers serves a legitimate role in facilitating the timely analysis and presentation of clinical-trial data for public consideration."

Amsterdam had recruited patients for the trial but was not included as an author; he protested at the time to his boss, department chair Dwight Evans. Amsterdam was prompted to file his current complaint with

the ORI after seeing allegations late last year that Evans had lent his name to an editorial (D. L. Evans and D. S. Charney *Biol. Psychiatry* **54**, 177–180; 2003) written by an STI writer who was being paid by GSK (the payment was not acknowledged in the publication). At the time, the university decided that the allegation of ghostwriting was unfounded.

Amsterdam's charges could prove awkward for the president of the University of Pennsylvania, Amy Gutmann, who is also the chair of US President Barack Obama's bioethics commission. In an 11 July letter to Obama, the



Project on Government Oversight (POGO), a watchdog group based in Washington DC that Amsterdam contacted while developing his complaint, called for Gutmann's ousting as chair. The letter takes issue with Gutmann's handling of the earlier ghostwriting allegations. "We do not understand how Dr. Gutmann can be a credible Chair of the Commission when she seems to ignore bioethical problems on her own campus," POGO's executive director, Danielle Brian, wrote.

The university said on 11 July that its School of Medicine will investigate the new allegations. The school's policy, adopted last year, states that medical researchers "are prohibited from allowing their professional presentations of any kind, oral or written, to be ghostwritten by any party, including Industry". The published paper acknowledged that GSK funded the study, but did not note that STI had been employed in the manuscript's preparation, or that three of the co-authors were GSK employees while the study was being conducted. The GSK authors are not included in Amsterdam's complaint.

The five authors whom Amsterdam accuses are Evans, Charles Nemeroff, now chairman of psychiatry at the University of Miami in Florida; Laszlo Gyulai, a psychiatrist at the University of Pennsylvania who has now retired; Gary Sachs, a psychiatrist at Massachusetts General Hospital in Boston; and Charles Bowden,

chairman of psychiatry at the University of Texas Health Science Center in San Antonio.

Evans and Gyulai did not respond to interview requests, but the university stated that "both Penn faculty members have been advised of the allegations in the complaint and while they believe them to be unfounded, have made clear to the University that they will fully cooperate with the investigation". Bowden says: "I provided input that was incorporated into the manuscript ... I never had any sense that the manuscript was 'ghostwritten'."

Sachs says he strongly agrees and that he "went physically from Boston to Philadelphia to draft the first draft" with Gyulai. The multi-site clinical trial was conducted in the mid-1990s and funded by GSK (SmithKline Beecham when funding was initiated). It compared Paxil — marketed as Seroxat outside the United States — the firm's new antidepressant, with imipramine, an older, cheaper, antidepressant, and with placebo in treating depression in people with bipolar disorder — a condition with a high suicide risk. Amsterdam alleges that the study: didn't enrol enough patients to come to definitive conclusions; made specious distinctions between subsets of subjects that allowed it to claim a positive result for Paxil in some patients; and played down the side effects of the drug. Nemeroff, the paper's first author, says that the data used withstood rigorous peer review in a process that sent the paper back to the authors for revisions several times. "Right in the abstract under 'results' we report that 'Differences in overall efficacy among the three groups were not statistically significant'," he says. "I don't know how much more straightforward we can be than that."

He adds that "with a 2011 magnifying glass, obviously one would have included in the published paper the use of an editorial assistant". Still, he says: "All [STI] did was help collate all the different authors' comments and help with references. We wrote the paper."

Paul Root Wolpe, a bioethicist at Emory University in Atlanta, Georgia, who reported to Evans and collaborated with Amsterdam while on the faculty of psychiatry at the University of Pennsylvania, says that the documents imply but do not prove that the manuscript was ghostwritten. But, he says, they indicate "a troubling level of control of pharma over the academic product".

Wolpe adds: "This is not an isolated case, but a systemic problem that needs a coordinated, systemic solution." ■



Japanese wheat farmers are unlikely to reap a radioactive harvest in future years.

## RADIATION MONITORING

# No fallout legacy for Japan's farms

*But the most contaminated soils need urgent clean-up.*

BY DAVID CYRANOSKI

After the Fukushima nuclear disaster spewed radiation across northern Japan in March, some feared that farming there would be shut down for years. But early studies of how the radiation has accumulated in plants and the soil now suggest that farmers in much of the region can go back to work.

Soon after the meltdown at Fukushima Dai-ichi, the government evacuated people living within 30 kilometres of the plant, and later imposed restrictions on agricultural products. Those measures are still in place, and the government has not yet announced a clear strategy for dealing with the contaminated areas. "People are panicking because there are no data," says plant radiophysiology expert Tomoko Nakanishi at the University of Tokyo.

Nakanishi is coordinating seven teams to study the impact of the disaster on soil, plants, animals, fisheries and forests for the next decade, measuring contamination levels and assessing the long-term threat. Their first results, to appear in the Japanese journal *Radioisotopes* in August, paint a surprisingly optimistic picture.

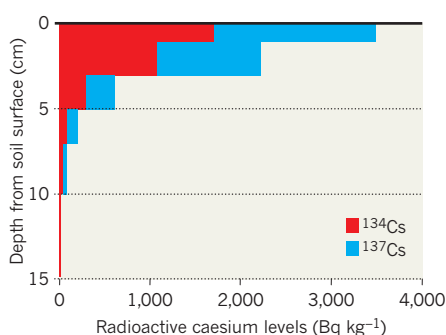
The scientists studied crops at a Tokyo research field, including cabbages and potatoes that were planted a few weeks after rains showered the field with radioisotopes from Fukushima. The crops were harvested on 16 May, and contained low levels of radiation — around 9 becquerels per kilogram (Bq kg<sup>-1</sup>; wet

weight), much lower than the 500 Bq kg<sup>-1</sup> safety limit for human consumption. Furthermore, most of the radiation had accumulated on the leaves and could be washed off, suggesting that the plants were not absorbing dangerous levels of radioisotopes directly from the soil.

The more highly exposed fields around Fukushima showed similar results, with most of the radiation in plants accumulated on their surfaces. Wheat leaves that were open at the time of the greatest fallout were heavily contaminated, with combined levels of caesium-134 and caesium-137 ranging from thousands to about 1 million Bq kg<sup>-1</sup>. But leaves that unfolded afterwards were largely free of contamination. Wheat ears from these plants contained 300–500 Bq kg<sup>-1</sup> — within the prescribed

## SKIN DEEP

Soil contamination is limited to the top few centimetres of fields around Fukushima.



radiation limit. "It's harvest time now and farmers are wondering what to do," says Nakanishi. "They can throw the current harvest away. But it is OK to plant again."

Despite this good news, the team's data also show that the radioisotopes seem to be stuck firmly to the soil, mainly in the top five centimetres (see 'Skin deep'), and are not being washed away by rain. This might prevent the radioisotopes from entering groundwater, but suggests that cleaning up the more radioactive public spaces in Fukushima prefecture will not be easy.

A separate group from Kobe University, led by radiation expert Tomoya Yamauchi, has found that soil radiation levels at four sites in Fukushima city, some 60 kilometres from the reactors, measured up to 47,000 Bq kg<sup>-1</sup> — surpassing the 10,000 Bq kg<sup>-1</sup> human exposure safety level set by the government. Yamauchi says that these areas, which are outside the current 30-kilometre evacuation zone, should be evacuated immediately.

In May, the agriculture ministry unveiled a ¥490-million (US\$6-million) initiative to develop clean-up techniques, including removing contaminated soil. But results from the tests won't come for months, and Nakanishi says that, in the meantime, the information gap is dangerous. Without data on the true depth of soil contamination, local schools are using large machines to scoop up the top 50 centimetres of soil — probably much more than is necessary — and leaving it as radioactive mounds in the corners of school playgrounds.

The agriculture ministry is also testing how well plants can clean the soil in highly contaminated areas, and several non-governmental organizations have followed suit with a campaign of sunflower planting. Nakanishi says that the effort is "nonsense", arguing that such phytoremediation would absorb only small amounts of radioisotopes. Chihiro Inoue, an expert in soil and groundwater remediation at Tohoku University, says that phytoremediation is worth testing, but warns that even if it works, "you're still left with the problem of how to dispose of the [radioactive] plants".

Burying the soil is expensive, however. Inoue says that the cost of cleaning up a school playground could be ¥50 million, and there are more than 100 schools in the affected areas, not to mention parks and other public places. Given that caesium-137, with its 30-year half-life, will be around for a while, soil burial sites would have to be monitored to make sure contaminated soil was not exposed by weather, he says.

Whatever happens, it needs to happen soon, Inoue says. People cannot rebuild their lives until the radiation risks are understood and a plan for reducing them is in place. "We can't wait much longer," he says. ■



WWW.NATURE.COM/JAPANQUAKE

SOURCE: T. NAKANISHI





# A spot of trouble

*By raising hell about newborn blood-spot screening, Twila Brase could jeopardize public-health programmes and derail research. The problem is, she has a point.*

**By Mary Carmichael**

**T**wila Brase was not always the kind of person who hands out politically charged propaganda in airports. On a first meeting at her modest office in a shopping plaza in St Paul, Minnesota, she seems more like the unassuming nurse she was back in 1995 — before she began her second life as a bioethical gadfly, and before she had started making YouTube videos that accuse her state of commandeering the DNA of children as “government property” through widespread newborn screening programmes. Her voice is quiet and level. It is difficult to write her off as a conspiracy theorist: she simply doesn’t sound like one, even when, 4.5 minutes into making the case against screening, she suggests that “some researchers” might be trying to convince the state to test day-old infants for genes linked to “a tendency towards violence”.

But Brase may have brought newborn screening and associated research in Minnesota to the point of crisis with her allegations. By tapping into ideological veins that run deep in the United States — wariness of government intrusion and fears about threats to privacy — she could influence the fates of many studies that seek to use human samples from state biobanks, not to mention the fates of thousands of children with rare diseases. Brase has a bent for the hyperbolic: her website features, among other touches, a photo of an infant in a shirt that reads, “Help! The Gov’t Has My DNA”. She also bends the truth at times to make her points. The claim that the state might test for “propensity to violence”, for example, is based on a self-published proposal by a student at the University of Connecticut School of Law in Hartford ([go.nature.com/uodhkh](http://go.nature.com/uodhkh)) — hardly an imminent programme.

But one argument gives her detractors pause. Parental consent for research on infant blood spots is handled poorly — if it is broached at all — in the laws of many US states. What little parents know about newborn screening often comes from a short brochure given to them just after labour and delivery, when

DARIN BACK/REDUX/EVEVINE



they're too distracted to process the contents.

Supporters of the screening tend to emphasize the medical benefits. "This is not about privacy or invading anyone's life," says Nancy Mendelsohn, a medical geneticist at Children's Hospitals and Clinics of Minnesota in Minneapolis. "These aren't things that we're doing to children. These are things that we're doing for children." But some state health departments are already rethinking their approach to informing parents and asking for consent.

### HEEL PRICKS FOR HEALTH

Among public-health professionals worldwide, newborn screening is generally viewed as one of the most successful innovations of the modern era. Within a few days of birth, infants are pricked at the heel, and their blood is tested for rare genetic and endocrine conditions that can be harmful or fatal if they are not caught early. Screening programmes began in the 1960s with a test for phenylketonuria, a disorder with effects on mental development that can be avoided through dietary restrictions. In the past decade, screening has expanded to encompass roughly 40 diseases. In most cases, DNA is not tested; rather, protein analysis screens for enzymes that might be affected by disease. Most developed countries have some form of newborn screening in place. The US programme alone identifies at least 3,400 children in need of treatment every year.

Beyond the public-health initiative, however, the United States and many other countries save 'blood-spot' samples on cards, and some give them to scientists for use in population-based studies, after stripping away identifying details. Advocates of biobanking view blood-spot repositories as a valuable scientific resource. The samples have been used to develop tests for debilitating and fatal disorders, such as severe combined immunodeficiency, and to ensure the accuracy of existing tests. They have also been used in epidemiological research — for instance, blood spots have helped scientists in Minnesota to examine

blood mercury levels and prenatal exposures to tobacco.

Because it is performed on tissue samples rather than on live human beings, such research generally does not require explicit informed consent. And parents are often uninformed. (Although some countries — such as the United Kingdom, Germany and the Netherlands — do have informed-consent policies for screening.) A 2009 survey conducted in part by Genetic Alliance, a research and health-care advocacy group in Washington DC, found that 62% of new mothers in the United States were not given any information about newborn screening, were not given

together by the Citizens' Council for Health Freedom (CCHF) in St Paul, an advocacy group that Brase leads. If the group loses, it could appeal. If at any point it wins, it could set a precedent for public-health officials and researchers across the country and around the world.

Brase became a privacy activist in the 1990s, during attempts at health-care reform by the Bill Clinton administration. She believed that government-imposed decisions on health care could affect people at their most vulnerable times. "I was a nurse, so I understood that patients often cannot protect themselves," she says. "I just said, 'I have to do

saying that parents could opt out of screening or storage, provided that they did so in writing. But that wasn't good enough for Brase. "They didn't say that there was an official form or tell people where they could find it," she says. So she began a second campaign, this time arguing that the newborn screening programme violated a genetic-privacy bill passed by the state in 2006. She won that too, although perhaps not in the way she wanted: in 2008, the legislature voted to exempt newborn screening from the bill. Brase's efforts had tied up the capitol for weeks.

She does not shoot for the subtle. Brase frequently name-checks the 1997 dystopian science-fiction film *Gattaca*, in which genetically 'inferior' people form a social underclass, and when she testified to the state House of Representatives in 2009, she placed two books in front of her: one about the US eugenics movement, the other about the Holocaust. She podcasts. She tweets. And every time she flies, she takes a stack of wallet-sized cards to hand out at the gate. "Protect your baby," they read, "Reclaim their DNA!"

Brase's rhetoric may be overblown, but in Texas, many share her concerns. Two years ago, an investigative journalist discovered that Texas had been shipping blood-spot cards to the US military, which was trying to build a national mitochondrial-DNA database for forensic identification. The state had also been bartering with private companies, trading blood spots for lab equipment. Worse, it had been trying to keep the initiatives under wraps: in an e-mail obtained by the *Texas Tribune*, one researcher argued against informing the public, saying that a press release would "only generate negative publicity". After a series of exposés and a lawsuit, Texas had to incinerate 5.3 million cards. The fallout continues: in May, the state legislature voted to change the research portion of its newborn screening programme from opt-out to opt-in.

The Minnesota Department of Health has never been accused of a Texas-style cover-up. And although many people within the department consider Brase something of an enemy of the state, they



Blood spots are used to screen for rare diseases, and sometimes in research.

enough information, or did not remember whether they had been given any.

Informed consent is central to Brase's campaign. This year, she worked with a Minnesota legislator to introduce a bill amendment that would change the state's entire screening programme — not just the research portion — from an opt-out model to an opt-in one. It also required the destruction of blood spots within hours of testing. Opponents said that the policy would result in the deaths of children and the shutdown of labs statewide — federal law requires that such samples be kept for two years for quality control — and the amendment was shelved. But Brase had another weapon. The state's Supreme Court is now considering a case on the same issues, filed by nine families brought

something." Brase took a break from nursing in 1995 to found the CCHF, and never went back.

Her attention turned to newborn screening when she was poring through an annual state appropriations bill one day in 2003. "I remember on page 80, I got to this language that said essentially that the health department would have the discretion to test every child for whatever conditions it wanted without having to come back to the legislature. And that was when I realized that newborn screening was not just newborn screening — it was genetic testing."

Brase became the bane of the state health department. She lobbied for better education of parents and got it: in the mid-2000s, Minnesota added a note to its newborn-screening brochure

rarely speak publicly about her. But Edward Ehlinger, Minnesota's commissioner of health, says that she has, in one sense, been helpful. "We think data privacy is incredibly important, and we also think individuals should know how information is going to be used. What Twila has done is to make sure we have those conversations," he says. Still, he adds, "as with any conversation, you do need to come to a place where you can move on."

More than a few of Brase's critics say that she could have an effect similar to that of Andrew Wakefield, the disgraced British doctor whose fraudulent research led millions of parents to believe in a link between vaccines and autism. Brase could cause parents to shun disease screening. She says that according to her figures, more Minnesota parents have declined newborn screening each year since 2003 — the parents of 156 children refused it last year. The more children go unscreened, the more likely it is that some with debilitating or fatal diseases will go untreated, says Mendelsohn. "With some of these disorders, if they're not caught quickly, the kids lose IQ points by the week."

#### RESEARCH PARALYSIS

Selling parents on the long-term benefits of research can be difficult. Many of the projects listed on the Minnesota Department of Health's website have not been written up and published, despite years of apparent work. The controversy has stymied others. Piero Rinaldo, a researcher into biochemical genetics at the Mayo Clinic in Rochester, Minnesota, had hoped to conduct pilot studies on several rare disorders to add to the state panel. A validated test could help doctors to treat the diseases earlier. Rinaldo says that he has taken all the required ethical precautions, but with the state "in paralysis," he has not been able to begin. "We're doing this because we want to save more lives," he says. "But [Brase] acts as if it's all an excuse for the government to build some inventory of imperfect children. It's like she has the idea that all science is bad."

Brase herself says almost as much. "I have a less glorified

sense of research than some people do," she says. "It seems like there's no final answer to some of the questions." She points to studies overturning previous recommendations on hormone-replacement therapy. For years, apparently well-founded advice urged women to take hormones, only to be overturned when later, better studies showed that they could be harmful.

To ensure that screening programmes aren't affected by attacks on the research, some experts

proposal that has gained traction among bioethicists is to educate parents earlier. "It's very clear that this shouldn't be done in the peripartum period," says Ellen Clayton, a bioethicist at Vanderbilt University in Nashville, Tennessee. The United Kingdom already has such a policy; its prenatal educational programmes begin with a leaflet given to parents in the third trimester of pregnancy. For that matter, Minnesota already provides information on newborn screening to obstetricians.



"I realized that newborn screening was not just newborn screening — it was genetic testing."

propose separating the consent processes — making screening opt-out and research opt-in, with more comprehensive information given to parents. "It seems to me you have to," says James Evans, a medical geneticist at the University of North Carolina in Chapel Hill, who has advised the government on bioethical issues. "If as many people don't participate in research, that is unfortunate. But when people don't participate in newborn screening, babies die." Michigan's public-health department, which has made a point of openly discussing newborn screening with the public, has settled on this approach. Another

But neither change would fully appease Brase, because the government would still hold children's blood samples, and thus potentially their genetic information. "Yes, the government has to follow the law, but nobody's in there watching. So the best way to make sure it never happens is to simply not get screened," she says. Confronted with the point that children with the targeted conditions could die if their parents opt out, she says that parents could have their infants tested privately instead. The right to make that choice should be theirs, she adds.

The debate is likely to become more ferocious — and more

complicated — as genome sequencing begins to enter medical practice. Population-level sequence data would be a gold mine for researchers. But many parents who have no problem with limited newborn screening might well feel uncomfortable having their children's entire genome sequences on file, no matter how strict the privacy protections.

Research that makes use of blood spots is likely to increase. The National Institutes of Health has established the Newborn Screening Translational Research Network, a group intended to facilitate data sharing and encourage more scientific work. One

of the many goals of the incipient network is to make it easier for scientists to access the millions of blood spots nationwide. The network's proponents recognize that it could engender controversy. In February, in the *American Journal of Public Health*, several members of the advisory committee wrote that parents are too poorly informed, and that "addressing concerns from stakeholders will be necessary for state-level adoption of national recommendations". (E. W. Rothwell *et al.* *Am. J. Public Health* doi:10.2105/AJPH.2010.200485; 2011).

This is something that Brase can agree with. "I really believe if you do not respect the rights of people, research is not going to be trusted in the future," she says. "Researchers will be looked at as people who want to stamp on your rights to get their grants and their fellowships and their chairs to prop themselves up."

There are more nuanced ways of putting it, but many of Brase's opponents concede the essence of her point. "If scientists want to be able to do science, they have to convince the public that it's a good thing to do, that there are protections in place, and that the practices are transparent," says Clayton. "Newborn screening cannot fail to do that." That will mean listening to objections, even if they come from the likes of Brase — otherwise, there might one day be many more like her. ■ **SEE EDITORIAL P.139**

**Mary Carmichael** is a freelance writer in Boston, Massachusetts.

DARIN BACK/REDUX/EVINE





# RISE OF THE TITANS

The sauropods were the biggest creatures ever to walk the planet. But the keys to their success emerged in their tiny ancestors.

BY FREDRIC HEEREN

From tail to snout, they stretched as long as four London double-decker buses parked end-to-end. The largest grew from 10-kilogram hatchlings to 100,000-kilogram adults. Their legs alone weighed several tonnes. No land creatures before or since have ever attained the size of the sauropod dinosaurs.

Those four-legged titans of the Jurassic and Cretaceous periods, 200 million–65 million years ago, had a suite of specializations that enabled them to reach such immense proportions. With long necks, wide-opening jaws and rake-like teeth, *Diplodocus*, *Brachiosaurus* and their ilk swept their heads through the tree-tops, consuming vast amounts of foliage without expending a lot of energy moving their massive legs. Adaptations of the pelvis and limbs created a frame sturdy enough to support their heft, and hollowed-out vertebrae and relatively small heads lightened the load. Their specialized bone development made it possible for juvenile sauropods to grow quickly, putting on several tonnes per year.

Palaeontologists have long thought that these anatomical novelties arose with the large sauropods — that a burst of evolutionary specializations coincided with the explosion in size. But a slew of discoveries in recent years reveals that many important changes first showed up long before, among the relatively puny forerunners of sauropods known as the early sauropodomorphs. Paul Barrett, a palaeontologist at the Natural History Museum in London, calls this group “the unsung members of the dino community”.

Walking upright on two legs, the early sauropodomorphs looked nothing like the lumbering beasts that came to dominate later. But these small creatures and their descendants

gradually acquired adaptations that changed how they ate, moved and breathed — in ways that would later enable sauropods to achieve their size (see ‘How to build a giant’).

“It is not that sauropods have these characters because they were gigantic,” says Diego Pol, a palaeontologist at the Egidio Feruglio Palaeontological Museum in Trelew, Argentina. “Instead, they achieved their gigantic size because they evolved from small-bodied ancestors that already had these features.”

## STAGE 1: STARTING SMALL

The discoveries have not come easily. Many of the relevant fossils were found in remote sites in the Southern Hemisphere, including Argentina and South Africa.

In 2006, palaeontologist Ricardo Martinez discovered a promising set of bones in the desert of northwestern Argentina. They emerged from rock that dated to the late Triassic Period, about 230 million years ago — a time when the first dinosaurs were starting to appear. He hauled the prize back to the Natural Sciences Museum in San Juan, then spent months freeing a lower jaw from the surrounding rock. Martinez found that the teeth

had coarse serrations along their edges, an adaptation for cutting through fibrous plant material. Other early dinosaurs had fine serrations, more suitable for slicing through flesh. This told Martinez that he had found a tiny predecessor of the great sauropods, one with

a relatively big skull like its carnivorous ancestors, but teeth more like those of an omnivore.

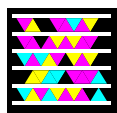
In 2009, Martinez and Oscar Alcober, also at the San Juan museum, described<sup>1</sup> the partial skeleton as the earliest and most primitive sauropodomorph yet found. Moving on two legs, the 1.6-metre-long animal had a body the size of a turkey, and a long tail. It weighed only 7–8 kilograms. Martinez called it *Panphagia protos*, meaning ‘first eater of everything’, to celebrate its step on the road from carnivory to herbivory.

Barrett lists “the greater reliance on vegetation rather than animal food” as one of the factors that “kick-started the increase in body size”. The advantage of herbivory is in the logistics of gathering food. Giant sauropods would never have been able to find and catch enough prey to fill their daily nutritional quota — which might have neared a tonne’s worth for the largest.

Traditional grazing could not have done the job either, says Martin Sander, a palaeontologist at the University of Bonn in Germany. Instead of continuously shifting locations, burning through energy as they hoisted their colossal legs, sauropods swung their heads back and forth, mowing efficiently through the foliage.

That kind of feeding required long necks, which would have been impossibly heavy if they were built with solid vertebrae. But large sauropods had vertebrae riddled with holes. These air-filled, or pneumatic, bones weighed only about 35% as much as solid ones, which helped the sauropods to carry necks up to 15 metres long, says Mathew Wedel, a palaeontologist at the Western University of Health Sciences in Pomona, California. Hollow areas within the pneumatic bones may have connected to air sacs in the body cavity that helped to blow air through the lungs and improved the breathing

**NATURE.COM**  
For an interactive  
tour of sauropod  
evolution, visit:  
[go.nature.com/c7z1ct](http://go.nature.com/c7z1ct)



Scan the tag above with the  
free app from **gettag.mobi**

# HOW TO BUILD A GIANT

The evolution of sauropods can be split into four stages, starting with tiny dinosaurs in the late Triassic (about 230 million years ago).

## STAGE 1: EARLY SAUROPODOMORPHS

Small, fleet, bipedal animals just 1–2.5 metres long that are among the oldest known dinosaurs.



## STAGE 2: PROSAUROPODS

Bipedal creatures that reached up to 10 metres long. Some had specializations for rapid bone growth.



## STAGE 3: NEAR SAUROPODS

Specialized prosauropods with adaptations that made their limbs and backbones sturdier.



## STAGE 4: SAUROPODS

Biggest land animals ever. *Brachiosaurus* (shown right), reached about 25 metres long, and fragmentary fossils hint at much larger species.

efficiency of the giants — features seen in modern birds. Without the extra volume provided by such air sacs, it would have been impossible for the sauropods to clear the stale air that filled their necks after each breath; their lungs were simply too small to do the job alone.

Pneumatic vertebrae would seem to be an adaptation related to giant size. But Wedel has found potential precursors in a small, early sauropodomorph named *Pantyraco*. Its neck vertebrae have pits that match the positions of the holes in the sauropod vertebrae<sup>2</sup>.

So how could the precursors of air sacs and pneumatic bones aid tiny dinosaurs? Researchers suspect that they increased the efficiency of oxygen exchange, possibly helping the ancestors of dinosaurs to out-compete their contemporaries during the late Permian and early Triassic periods (260 million–240 million years ago), when atmospheric oxygen concentrations were much lower than they are today<sup>3</sup>.

## STAGE 2: ADDING TONNES PER YEAR

The earliest sauropodomorphs were small, fast, and mostly moved on two legs. They could rely on speed to evade predators. But the next stage in evolution took the creatures a step up in size, to between 2 and 10 metres long.

The oldest known fossils of these ‘core prosauropods’ date from the start of the Jurassic period, about 200 million years ago. These creatures had longer necks and torsos, with larger bodies and relatively shorter legs than their predecessors. That made prosauropods less nimble, but their size helped to keep them safe.

That defence took its most extreme form with the later sauropods. “Adult sauropods presumably were almost immune from predation because of their body mass being an order of magnitude greater than that of the largest predators,” says Sander. “Their sheer volume made it difficult for an attacker to place an effective bite.”

If sauropods grew slowly like most reptiles, each one could have taken more than one hundred years to reach full size. But that would have left the smaller juveniles vulnerable for decades. Instead, evidence is emerging that these dinosaurs grew much faster than modern reptiles.

The key innovation was fibrolamellar

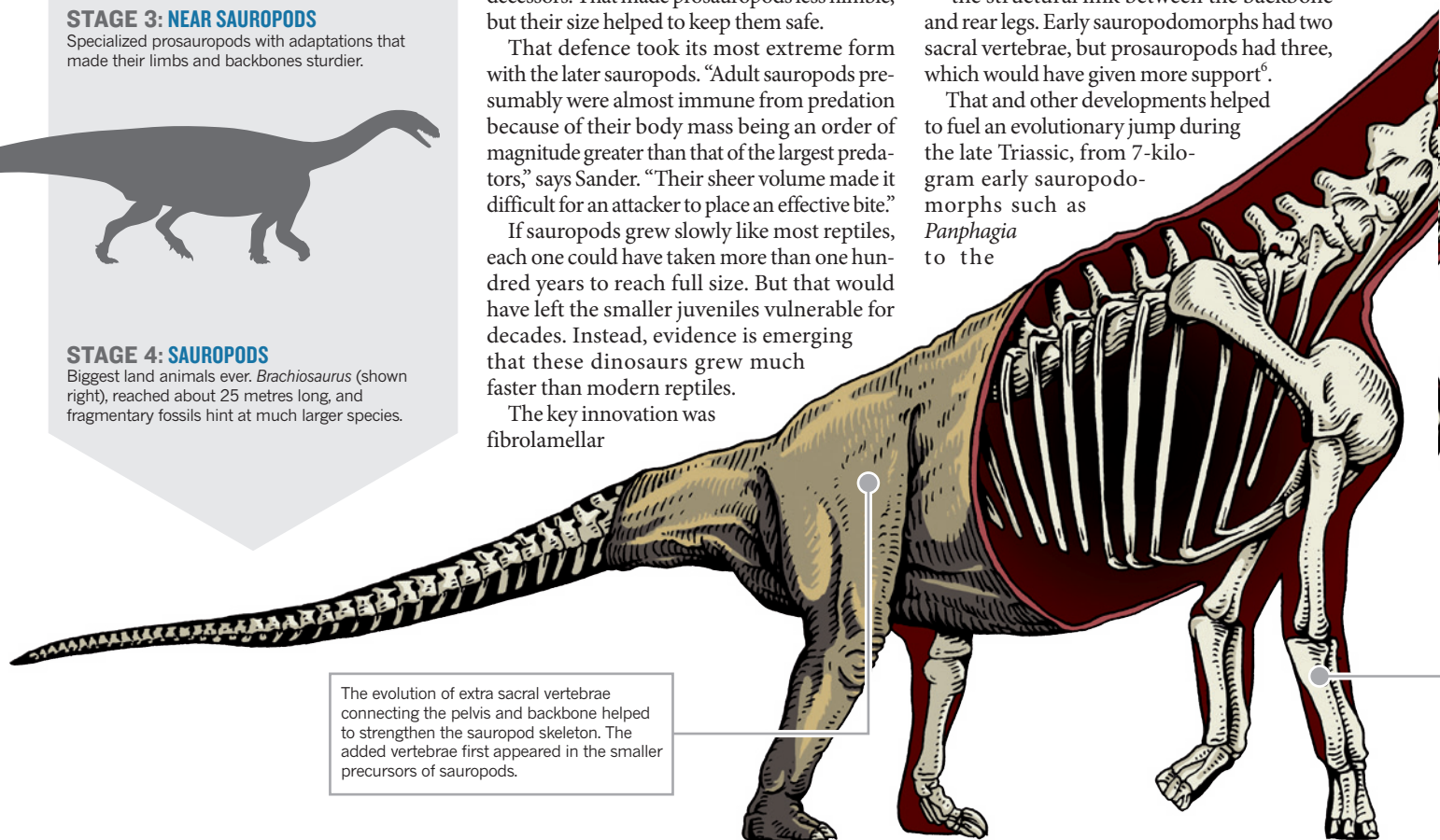
bone, which develops in two stages, says Sander. “A scaffold of bone is thrown up very quickly, making the bone grow in thickness by about one tenth of a millimetre per day, which then is filled in more gradually.” Over the past decade, Sander and other researchers have analysed the structure of fossilized bone and documented the presence of fibrolamellar bone in sauropods. He estimates that the animals could grown by a few tonnes per year.

But the origins of this trait appeared long before the giant sauropods. In 2005, Sander and one of his graduate students, Nicole Klein, reported<sup>4</sup> signs of fibrolamellar bone in *Plateosaurus*, a core prosauropod that lived during the late Triassic and reached only about 10 metres long. By studying bones from more than 40 *Plateosaurus* individuals in Germany, Klein and Sander found that some reached full size in as little as 12 years.

Growth that fast is more characteristic of a warm-blooded animal than a cold-blooded one, and some dinosaurs might have had elevated body temperatures. Robert Eagle, a geochemist at the California Institute of Technology in Pasadena, and his colleagues reported<sup>5</sup> last month that two giant sauropods, *Brachiosaurus* and *Camarasaurus*, had body temperatures 5–12 °C higher than those of modern alligators.

*Plateosaurus* and other prosauropods showed further anatomical developments that later helped their descendants to achieve massive size. For example, they had a beefed-up sacrum — the structural link between the backbone and rear legs. Early sauropodomorphs had two sacral vertebrae, but prosauropods had three, which would have given more support<sup>6</sup>.

That and other developments helped to fuel an evolutionary jump during the late Triassic, from 7-kilo-gram early sauropodomorphs such as *Panphagia* to the



The evolution of extra sacral vertebrae connecting the pelvis and backbone helped to strengthen the sauropod skeleton. The added vertebrae first appeared in the smaller precursors of sauropods.



4,000-kilogram *Plateosaurus*. “The dramatic size increase observed along the first 25 million years of sauropodomorph history was the fastest one in the history of life,” says Martin Ezcurra, a palaeontologist at the Bernardino Rivadavia Argentinian Museum of Natural Sciences in Buenos Aires.

### STAGE 3: EDGE OF GREATNESS

Some of the most recent fossil discoveries fall into a third chapter of sauropodomorph evolution: creatures that could be called near-sauropods. Adam Yates, a palaeontologist at the University of Witwatersrand in Johannesburg, came to South Africa hoping to find fossils from this stage that could reveal how sauropodomorphs became quadrupedal.

He and a student hit the jackpot on a hill called Spion Kop. “Bone was piled upon bone,” says Yates. “Given that we were finding so much of the skeleton, including parts of the

small, fragile skull, we knew this was a major find.”

Last year, Yates and his colleagues named the new species *Aardonyx celestae*<sup>7</sup>. From the lower jaw of *Aardonyx*, Yates could tell that it did not have fleshy cheeks that limited how far the jaw could gape open. Instead of taking small bites and chewing as its older relatives did, *Aardonyx* could have opened its jaws wider and grabbed big mouthfuls, gulping them down whole.

That adaptation enabled the development of extremely long necks among sauropods, because it did away with the need for big jaw muscles and massive heads. “The long neck was only possible because they did not chew,” says Sander.

*Aardonyx* was bipedal, but it had acquired some leg features that show up in quadrupedal sauropods, with their lumbering gait. Matthew Bonnan, a palaeobiologist at Western Illinois University in Macomb and a co-author on the *Aardonyx* paper, says that the creature’s thigh bones were longer than the bones of its lower legs, unlike earlier sauropodomorphs, in which these bones were about the same size. “This alone suggests animals that were built not for speed but for support,” says Bonnan.

The forelimbs of *Aardonyx* also showed quadruped-like adaptations. In true sauropods, the two long bones of the forearm interlocked in a way that made the front limbs sturdier. *Aardonyx* showed an earlier stage of this interlocking forearm, connected to a hand that could grasp. Before the discovery of *Aardonyx* and a related species called *Melanorosaurus*, Bonnan had hypothesized that such locking would produce an evolutionary chain reaction that also altered the hands in ways more suited to walking. He suggested that these adaptations would come in an “integrated functional suite” of shifting bones, which he expected to see first in a full-blown, quadrupedal sauropod. This hypothesis, he says, was “smashed” by the features of the bipedal *Aardonyx*.

This year, Pol described<sup>8</sup> another near-sauropod that demolished expectations. The early Jurassic dinosaur, *Leoneerasaurus taquetrensis*, was just 2.5 metres long and walked on two legs. But it had four sacral vertebrae. Just last year, Yates had written<sup>7</sup> that four sacral vertebrae were diagnostic of the four-footed posture.

Pol also found that *Leoneerasaurus* had spoon-shaped, forward-leaning front teeth for raking in vegetation — much like later sauropods. This 2.5-metre animal with many sauropod traits is helping to build a new picture of sauropod evolution that, Pol says, “has turned upside down the previous ideas”.

Researchers note that *Leoneerasaurus* and other known sauropodomorphs were not the ancestors of sauropods. Because the fossil record is so spotty, it is usually impossible to identify direct ancestors. But the prosauropods

and near-sauropods of the Jurassic preserve information about adaptations that appeared among the unknown ancestors of sauropods.

### STAGE 4: ON ALL FOURS

Many late Triassic and early Jurassic sauropodomorphs could walk on two legs or four, as needed. But in 2008, Ronan Allain, a palaeontologist at the National Museum of Natural History in Paris, and Najat Aquesbi, a palaeontologist at Mohammed V University in Rabat, described an animal from the later part of the early Jurassic that seemed committed to four<sup>9</sup>.

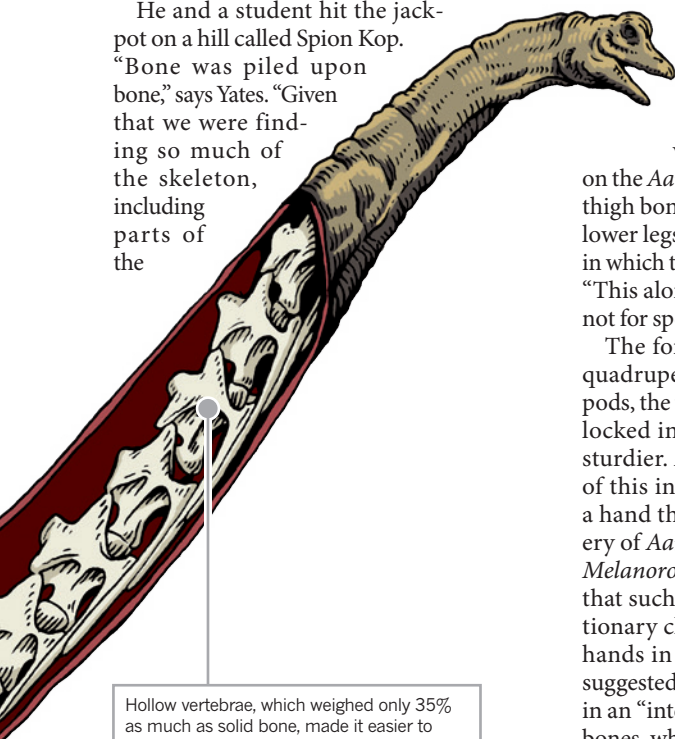
“*Tazoudasaurus* could be considered the oldest known ‘true sauropod,’” says Allain. Unlike its forebears, which had long fingers that could grasp, this 9-metre-long animal had a stubby hand suited to bearing weight. Allain placed *Tazoudasaurus* into a new group of sauropods that he named Gravisauria, or ‘heavy lizards’.

Heaviness is relative, and the most massive sauropods did not arise for another 90 million years. By the Cretaceous, fossils hint that some sauropods reached lengths of 40 metres and approached body masses of 100 tonnes. Yet in comparison with the changes that occurred during the early stages of sauropodomorph evolution, these later developments were relatively minor tweaks to a body plan that had emerged earlier.

The sauropod story shows the importance of pre-adaptations — traits that are neutral or serve some purpose but later become co-opted to fill a new function. Such traits constrain the future evolutionary pathways of a lineage, but with hindsight they can seem fortuitous for something that researchers consider an important attribute, such as gigantism. “The evolution of sauropods does look like kind of a crapshoot in which everything fell into place,” says Wedel. “Sauropods seem to have somehow gotten the evolutionary Wonka ticket of all the features that they needed to grow big.” ■ [SEE BOOKS AND ARTS P.172](#)

**Fredric Heeren** is a freelance writer in Olathe, Kansas.

1. Martinez, R. N. & Alcober, O. A. *PLoS ONE* **4**, e4397 (2009).
2. Yates, A. M., Wedel, M. J. & Bonnan, M. F. *Acta Palaeontol. Pol.* doi:10.4202/app.2010.0075 (2011).
3. Berner, R. A. *Geochim. Cosmochim. Acta* **69**, 3211–3217 (2005).
4. Sander, P. M. & Klein, N. *Science* **310**, 1800–1802 (2005).
5. Eagle, R. A. *et al. Science* doi:10.1126/science.1206196 (2011).
6. Rauhut, O. W. M., Fechner, R., Remes, K. & Reis, K. in *Biology of the Sauropod Dinosaurs: Understanding the Life of Giants* (eds Klein, N., Remes, K., Gee, C. T. & Sander, P.M.) 119–149 (Indiana Univ. Press, 2011).
7. Yates, A. M., Bonnan, M. F., Neveling, J., Chinsamy, A. & Blackbeard, M. G. *Proc. R. Soc. B* **277**, 787–794 (2010).
8. Pol, D., Garrido, A. & Cerda, I. A. *PLoS ONE* **6**, e14572 (2011).
9. Allain, R. & Aquesbi, N. *Geodiversitas* **30**, 345–424 (2008).



Hollow vertebrae, which weighed only 35% as much as solid bone, made it easier to support extremely long necks and tails. Precursors of the hollow bones appeared among the early sauropodomorphs.

Interlocking bones in the forelimbs increased stability and reduced flexibility. Some of the bipedal prosauropods showed early stages of the interlocking forelimb bones.





# COMMENT

**HISTORY** Mathematics can have huge practical utility, but you can't force it **p.166**

**EVOLUTION** A paean to the wonderful workings of feathers **p.170**

**EXHIBITION** Giant sauropods stalk American Museum of Natural History **p.172**

**FILM** Herbert Terrace, Nim Chimpsky researcher, in conversation **p.173**

CLOCKWISE FROM TOP LEFT: O. STREWE/LOVELY PLANET IMAGES; A. DISSANAYAKE/LOVELY PLANET IMAGES; K. COOLE/LOVELY PLANET IMAGES; J. HAGLUND/LOVELY PLANET IMAGES; P. ADAMS/GETTY IMAGES; K. CALVO VIA AP IMAGES; T. VOETEN/PANOS; R. TANSON/LOVELY PLANET IMAGES; J. HAGLUND/LOVELY PLANET IMAGES; P. ADAMS/GETTY IMAGES



## Genomics for the world

Medical genomics has focused almost entirely on those of European descent. Other ethnic groups must be studied to ensure that more people benefit, say **Carlos D. Bustamante, Esteban González Burchard and Francisco M. De La Vega.**

**I**n the past decade, researchers have dramatically improved our understanding of the genetic basis of complex chronic diseases, such as Alzheimer's disease and type 2 diabetes, through more than 1,000 genome-wide association studies (GWAS). These scan the genomes of thousands of people for known genetic variants, to find out which are associated with a particular condition.

Yet the findings from such studies are likely to have less relevance than was

previously thought for the world's population as a whole. Ninety-six per cent of

### SUMMARY

- Those most in need must not be the last to benefit from genetic research
- Reviewers and granting bodies must demand racial and ethnic diversity in genome studies
- Global genomics needs the financial support of governments and non-profits

subjects included in the GWAS conducted so far are people of European descent<sup>1</sup> (see 'Sampling bias'). And a recent *Nature* survey suggests that this bias is likely to persist in the upcoming efforts to sequence people's entire genomes<sup>2</sup>.

Geneticists worldwide must investigate a much broader ensemble of populations, including racial and ethnic minorities. If we do not, a biased picture will emerge of which variants are important, and genomic medicine will largely benefit a privileged few. ►



► Since the 1970s, geneticists have known that most of the genetic variance between individuals stems from differences in DNA sequence (genetic variants). Of the millions of small sequence differences identified worldwide, some are ‘common variants’ — that is, they are found in more than 5% of people in many populations, and some of these occur widely, in people of different geographical and ethnic origins. Indeed, most GWAS have sought to find common variants associated with disease, in the hope that discoveries in one population will generalize to others.

GWAS have unearthed clear associations between common variants and many common diseases. But depending on the condition in question, these explain only between 5% and 50% of the diseases’ heritability. Many of the genetic factors thought to be responsible are still ‘missing’.

This suggests that ‘rare’ genetic variants (those that occur in less than 5% of the world’s population but which comprise the bulk of genetic variants) may be disproportionately important<sup>3</sup> — both in determining a person’s risk of getting a complex disease and in predicting their response to a particular drug<sup>4</sup>. Rare variants tend to be population specific<sup>5</sup> (see ‘Comparing the uncomparable’). So if they do play a key part in disease, the lack of diversity in genetic studies will be severely skewing our understanding of which are important.

Several researchers have begun to assess the ability to generalize GWAS discoveries between different populations. Preliminary results suggest that findings from one population may not always easily translate to the rest of the world, although many more and larger comparisons are needed.

For example, in people with Native South American ancestry, a particular variant of a protein that transports cholesterol into cells is common and is strongly associated with low levels of high-density lipoprotein cholesterol, obesity and type 2 diabetes. European, Asian and African populations do not have this variant<sup>6</sup>.

Conversely, in dozens of studies in European populations, researchers have found 19 common single-nucleotide changes that are strongly associated with type 2 diabetes. In a further study of 6,000 people including European Americans, African Americans, Latinos, Japanese Americans and Native Hawaiians, 13 of these polymorphisms continue to be strongly associated with the disease<sup>7</sup>. Yet 5 of the 19 variants seem to have different effects in the different ethnic groups, and the role of one variant is unclear.

#### AVOIDING GENERALIZATIONS

There are several reasons why findings in one population might not generalize to another. Disease-associated versions (or alleles) of a

gene may vary substantially in frequency in different ethnic groups. Also, GWAS identify genetic markers associated with a particular trait, not the mutations causing the disease. If a given marker is linked to a mixture of common and rare causal alleles<sup>8</sup>, some of the rare ones are likely to differ in frequency in different populations, or even be completely absent in some<sup>5</sup>.

The degree to which common genetic markers are linked to underlying causal mutations will also vary depending on the population being studied. For example, African populations are generally more genetically

**“Replicating an association study in a different ethnic group is often one-tenth the original cost.”**

diverse than European, Asian or indigenous American populations, so one might expect to see weaker associations between markers and mutations in African and African-diaspora populations (such as African-Americans).

As well as genes and the environment differing between populations, the gene-environment interactions can vary, and these could significantly change a person’s likelihood of developing a disease.

Already there is evidence that measures of genetic ancestry can improve clinical care for people of mixed race. For example, physicians assessing the effects of lung disease compare measures of lung function (obtained by having the patient breathe into a spirometer) to a reference standard for healthy people of the same gender and racial group. Doctors make more accurate diagnoses when they use patients’ actual genetic ancestry to make comparisons, instead of self-selected or inferred categorizations of race or ethnicity<sup>9</sup>.

Likewise, researchers this year showed that Native American ancestry is associated with a greater risk of childhood acute

lymphoblastic leukaemia returning after a remission, and that children with more than 10% Native American ancestry need an additional round of chemotherapy to respond to the treatment<sup>10</sup>.

So why are geneticists reluctant to undertake studies of people with diverse ancestries?

A key concern is that as more populations are included in a study, it becomes increasingly likely that a sequence variant will be associated with disease because of differences in race or ethnicity between cases and controls, rather than because of differences in people’s disease status. A lack of appropriate methods or access to the right data may make it difficult for many investigators to control for this. Certainly, it is hard to both collect samples from tens or hundreds of thousands of patients, and balance costs, statistical power and project deadlines with broad ethnic representation.

Such challenges, however, do not justify restricting the beneficiaries of medical genomic research to a small subset of humanity. Population-based studies must be carried out on a global scale. This means giving incentives to researchers in developed countries to increase the representation of minority populations in their studies and — crucially — empowering investigators in the developing world to undertake genomics research themselves.

#### COST EFFECTIVE

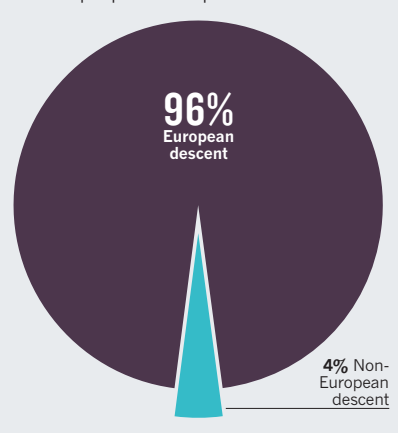
The ‘missing heritability problem’ has led many to become dismissive of GWAS. A danger of this GWAS fatigue is that it deters others from applying the approach to populations where it is likely to yield excellent results. GWAS has proved most successful in relatively small homogeneous populations — in Finland, Iceland and Costa Rica, say, where people generally stay put. Large families and limited migration are common among populations in Latin America, Africa and South Asia — suggesting that new and important associations between diseases and regionally common genetic variants may be found easily in these groups.

Moreover, large-scale GWAS are a feasible option for many research groups worldwide given that it now costs less than US\$250 to obtain genetic data across millions of markers per person. (Whole-genome sequencing costs about 20 times more.) In fact, replicating an association study in a different ethnic group is often one-tenth the original cost. So at the very least, associations found in Europeans should be investigated in other ethnic groups.

Key to the success of global-scale GWAS are extensive and accurate catalogues of human genomic variation. The 1,000 Genomes Project is an excellent first step to providing a reference resource for researchers

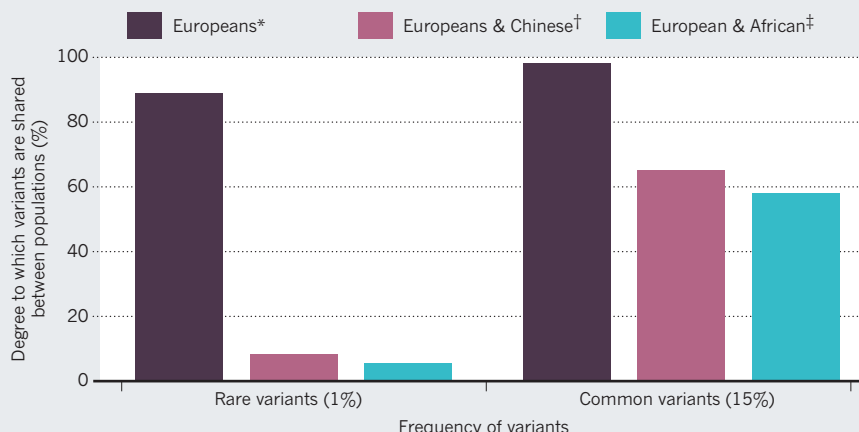
#### SAMPLING BIAS

Most genome-wide association studies have been of people of European descent.



## COMPARING THE UNCOMPARABLE

The rarer a genetic variant is within a population, the less likely it is to be found in all ethnic groups. One hundred people were sampled from each population.



\*Comparison of individuals of European descent in Utah and in Tuscany, Italy. † Han Chinese individuals from Beijing compared with Utah sample ‡ Yoruba individuals from Ibadan, Nigeria, compared with Utah sample.

worldwide. It aims to catalogue genetic variants occurring in more than 1% of various populations throughout the world — including mixed-race North and South Americans as well as diverse populations from Africa, Europe, the Far East and South Asia.

Boosting genomic studies globally will also require initiatives that foster collaboration between countries, and enable the transfer of funding and technology beyond China, the United States and the European Union. The Human Heredity and Health in Africa Initiative, for example, is empowering local researchers. Supported by the US National Institutes of Health (NIH) and the UK Wellcome Trust in London, the initiative is the first major attempt to help African investigators study the genomic and environmental determinants of common diseases in African populations.

The private and philanthropic sector can also play an important part. The Slim Initiative for Genomic Medicine, a collaboration between the Mexican National Human Genome Research Institute and the Broad Institute of the United States, was launched last year and is enabling researchers to study type 2 diabetes and cancer in Latin American populations. This project is funded by the charitable foundation of the Mexican business magnate and philanthropist, Carlos Slim Helú.

For these efforts to be successful, researchers and physicians in the participating developing countries cannot simply provide samples. In addition, local expertise, resources and technology centres must be developed so that local populations benefit directly from home-grown research — as at the BGI (formerly the Beijing Genomics Institute) in Shenzhen,

China. Local researchers will often better understand the history of local populations, such as whether they have recently switched from a rural to an urban lifestyle, and so will have deeper insights into likely environmental effects.

At the same time, medical geneticists working in wealthy nations must include their own minority and immigrant populations in their studies and develop the tools needed to compare results between populations. Many 'exome' sequencing projects (which sequence only the coding regions of the genome) are moving in this direction. For example, the US National Heart, Lung and Blood Institute aims to sequence the exomes of 7,000 people, roughly half of whom are African-Americans, to identify variants associated with cardiovascular and lung diseases.

### PIECING TOGETHER THE MOSAIC

To make medical genomics truly global, geneticists need new statistical methods to dissect the contribution of genetic, socio-cultural and environmental factors to both chronic and infectious disease.

Currently, 'ancestry metrics' are used to correct for the effect of shared ancestry on the results of association studies. But these methods do not work very well for groups whose genomes are a mosaic of fragments drawn from many different populations. Reference data from the relevant ancestral populations, including historically marginalized populations such as native Americans and Australian Aborigines, will help geneticists to separate spurious from real associations. And such understudied populations

must be properly on board for this to happen. Researchers should gauge local values and concerns, and invest time and money into education and outreach to explain to the people they intend to study, as well as to the general public, why studying global (and local) health is so important.

The Center for Research on Genomics and Global Health headed by Charles Rotimi at the US National Human Genome Research Institute in Bethesda, Maryland, is beginning to gather the data and formulate the methods needed to understand the complex interplay that creates health disparities among ethnic groups for diseases such as type 2 diabetes, hypertension and obesity in Africa and elsewhere. Meanwhile, the Slim, the Wellcome Trust and the Bill & Melinda Gates Foundation based in Seattle, Washington, have begun to support research in understudied populations. Ultimately, however, global genomics needs the financial support of governments.

One way to encourage researchers to branch out may be for peer reviewers and granting bodies to stress the importance of racial and ethnic diversity in medical genetic studies. The NIH mandated the inclusion of diverse subjects in 1985. In the 26 years since, just 7% of GWAS have included minorities — perhaps because being more inclusive doesn't win points for grant applicants.

It is tempting to focus on populations that are motivated, organized, medically compliant and otherwise easy to study. But by failing to develop resources, methodologies and incentives for underserved people, we risk perpetuating the health disparities that plague the medical system. Those most in need must not be the last to receive the benefits of genetic research. ■

**Carlos D. Bustamante and Francisco M. De La Vega** are in the Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA. **Esteban G. Burchard** is at the University of California, San Francisco, San Francisco, California 94143, USA. e-mail: cdbustam@stanford.edu

1. Need, A. C. & Goldstein, D. B. *Trends Genet.* **25**, 489–494 (2009).
2. *Nature* **467**, 1026–1027 (2010).
3. Manolio, T. A. et al. *Nature* **461**, 747–753 (2009).
4. McClellan, J. & King, M. C. *Cell* **141**, 210–217 (2010).
5. Gravel, S. et al. *Proc. Natl Acad. USA* doi: 10.1073/pnas.1019276108 (2011).
6. Acuna-Alonso, V. et al. *Hum. Mol. Genet.* **19**, 2877–2885 (2010).
7. Waters, K. M. et al. *PLoS Genet.* **6**, e1001078 (2010).
8. Adeyemo, A. & Rotimi, C. *Public Health Genomics* **13**, 72–79 (2010).
9. Kumar, R. et al. *N. Engl. J. Med.* **363**, 321–330 (2010).
10. Yang, J. J. et al. *Nature Genet.* **43**, 237–241 (2011).

The authors declare competing financial interests: for details see [go.nature.com/gvtxn4](http://go.nature.com/gvtxn4)





# The unplanned impact of mathematics

**Peter Rowlett** introduces seven little-known tales illustrating that theoretical work may lead to practical applications, but it can't be forced and it can take centuries.

As a child, I read a joke about someone who invented the electric plug and had to wait for the invention of a socket to put it in. Who would invent something so useful without knowing what purpose it would serve? Mathematics often displays this astonishing quality. Trying to solve real-world problems, researchers often discover that the tools they need were developed years, decades or even centuries earlier by mathematicians with no prospect of, or care for, applicability. And the toolbox is vast, because, once a mathematical result is proven to the satisfaction of the discipline, it doesn't need to be re-evaluated in the light of new evidence or refuted, unless

it contains a mistake. If it was true for Archimedes, then it is true today.

The mathematician develops topics that no one else can see any point in pursuing, or pushes ideas far into the abstract, well beyond where others would stop. Chatting with a colleague over tea about a set of problems that ask for the minimum number of stationary guards needed to keep under observation every point in an art gallery, I outlined the basic mathematics, noting that it only works on a two-dimensional floor plan and breaks down in three-dimensional situations, such as when the art gallery contains a mezzanine. "Ah," he said, "but if we move to 5D we can adapt..." This extension and abstraction

without apparent direction or purpose is fundamental to the discipline. Applicability is not the reason we work, and plenty that is not applicable contributes to the beauty and magnificence of our subject.

There has been pressure in recent years for researchers to predict the impact of their work before it is undertaken. Alan Thorpe, then chair of Research Councils UK, was quoted by *Times Higher Education* (22 October 2009) as saying: "We have to demonstrate to the taxpayer that this is an investment, and we do want researchers to think about what the impact of their work will be." The US National Science Foundation is similarly focused on broader

ILLUSTRATIONS BY DAVID PARKINSON

impacts of research proposals (see *Nature* **465**, 416–418; 2010). However, predicting impact is extremely problematic. The latest *International Review of Mathematical Sciences* (Engineering and Physical Sciences Research Council; 2010), an independent assessment of the quality and impact of UK research, warned that even the most theoretical mathematical ideas “can be useful or enlightening in unexpected ways, sometimes several decades after their appearance”.

There is no way to guarantee in advance what pure mathematics will later find application. We can only let the process of curiosity and abstraction take place, let mathematicians obsessively take results to their logical extremes, leaving relevance far behind, and wait to see which topics turn out to be extremely useful. If not, when the challenges of the future arrive, we won't have the right piece of seemingly pointless mathematics to hand.

To illustrate this, I asked members of the British Society for the History of Mathematics (including myself) for unsung stories of the unplanned impact of mathematics (beyond the use of number theory in modern cryptography, or that the mathematics to operate a computer existed when one was built, or that imaginary numbers became essential to the complex calculations that fly aeroplanes). Here follow seven; for more, see [www.bshh.org](http://www.bshh.org). **Peter Rowlett**

## MARK MCCARTNEY & TONY MANN

### From quaternions to Lara Croft

*University of Ulster, Newtownabbey, UK; University of Greenwich, London*

Famously, the idea of quaternions came to the Irish mathematician William Rowan Hamilton on 16 October 1843 as he was walking over Brougham Bridge, Dublin. He marked the moment by carving the equations into the stonework of the bridge. Hamilton had been seeking a way to extend the complex-number system into three dimensions: his insight on the bridge was that it was necessary instead to move to four dimensions to obtain a consistent number system. Whereas complex numbers take the form  $a + ib$ , where  $a$  and  $b$  are real numbers and  $i$  is the square root of  $-1$ , quaternions have the form  $a + bi + cj + dk$ , where the rules are  $i^2 = j^2 = k^2 = ijk = -1$ .

Hamilton spent the rest of his life promoting the use of quaternions, as mathematics both elegant in its own right and useful for

solving problems in geometry, mechanics and optics. After his death the torch was carried by Peter Guthrie Tait (1831–1901), professor of natural philosophy at the University of Edinburgh. William Thomson (Lord Kelvin) wrote of Tait: “We have had a thirty-eight-year war over quaternions.” Thomson agreed with Tait that they would use quaternions in their important joint book the *Treatise on Natural Philosophy* (1867) wherever they were useful. However, their complete absence from the final manuscript shows that Thomson was not persuaded of their value.

By the close of the nineteenth century, vector calculus had eclipsed quaternions, and mathematicians in the twentieth century generally followed Kelvin rather than Tait, regarding quaternions as a beautiful, but sadly impractical, historical footnote.

So it was a surprise when a colleague who teaches computer-games development asked which mathematics module students should take to learn about quaternions. It turns out that they are particularly valuable for calculations involving three-dimensional rotations, where they have various advantages over matrix methods. This makes them indispensable in robotics and computer vision, and in ever-faster graphics programming.

Tait would no doubt be happy to have finally won his ‘war’ with Kelvin. And Hamilton's expectation that his discovery would be of great benefit has been realized, after 150 years, in gaming, an industry estimated to be worth more than US\$100 billion worldwide.



## GRAHAM HOARE

### From geometry to the Big Bang

*Correspondence editor, Mathematics Today*

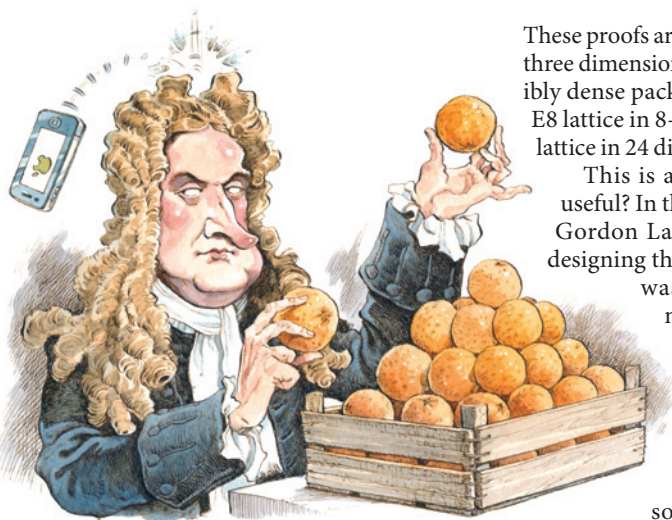
In 1907, Albert Einstein's formulation of the equivalence principle was a key step in the development of the general theory of relativity. His idea, that the effects of acceleration are indistinguishable from the effects of a uniform gravitational field, depends on the equivalence between gravitational mass and inertial mass. Einstein's essential insight was that gravity manifests itself in the form of space-time curvature; gravity is no longer regarded as a force. How matter curves the surrounding space-time is expressed by Einstein's field equations. He published his general theory in 1915; its origins can be traced back to the middle of the previous century.

In his brilliant Habilitation lecture of 1854, Bernhard Riemann introduced the principal ideas of modern differential geometry —  $n$ -dimensional spaces, metrics and curvature, and the way in which curvature controls the geometric properties of space — by inventing the concept of a manifold. Manifolds are essentially generalizations of shapes, such as the surface of a sphere or a torus, on which one can do calculus. Riemann went far beyond the conceptual frameworks of Euclidean and non-Euclidean geometry. He foresaw that his manifolds could be models of the physical world.

The tools developed to apply Riemannian geometry to physics were initially the work of Gregorio Ricci-Curbastro, beginning in 1892 and extended with his student Tullio Levi-Civita. In 1912, Einstein enlisted the help of his friend, the mathematician Marcel Grossman, to use this ‘tensor calculus’ to articulate his deep physical insights in mathematical form. He employed Riemann manifolds in four dimensions: three for space and one for time (space-time).

It was the custom at the time to assume that the Universe is static. But Einstein soon found that his field equations when applied to the whole Universe did not have any static solutions. In 1917, to make a static Universe possible, Einstein added the cosmological constant to his original field equations. Reasons for believing in an explosive origin to the Universe, the Big Bang, were put forward by Aleksander Friedmann in his 1922 study of Einstein's field equations in a cosmological context. Grudgingly accepting the irrefutable evidence of the expansion of the Universe, Einstein deleted the constant in 1931, referring to it as “the biggest blunder” of his life.





EDMUND HARRISS

## From oranges to modems

University of Arkansas, Fayetteville

In 1998, mathematics was suddenly in the news. Thomas Hales of the University of Pittsburgh, Pennsylvania, had proved the Kepler conjecture, showing that the way greengrocers stack oranges is the most efficient way to pack spheres. A problem that had been open since 1611 was finally solved! On the television a greengrocer said: "I think that it's a waste of time and taxpayers' money." I have been mentally arguing with that greengrocer ever since: today the mathematics of sphere packing enables modern communication, being at the heart of the study of channel coding and error-correction codes.

In 1611, Johannes Kepler suggested that the greengrocer's stacking was the most efficient, but he was not able to give a proof. It turned out to be a very difficult problem. Even the simpler question of the best way to pack circles was only proved in 1940 by László Fejes Tóth. Also in the seventeenth century, Isaac Newton and David Gregory argued over the kissing problem: how many spheres can touch a given sphere with no overlaps? In two dimensions it is easy to prove that the answer is 6. Newton thought that 12 was the maximum in 3 dimensions. It is, but only in 1953 did Kurt Schütte and Bartel van der Waerden give a proof.

The kissing number in 4 dimensions was proved to be 24 by Oleg Musin in 2003. In 5 dimensions we can say only that it lies between 40 and 44. Yet we do know that the answer in 8 dimensions is 240, proved back in 1979 by Andrew Odlyzko of the University of Minnesota, Minneapolis. The same paper had an even stranger result: the answer in 24 dimensions is 196,560.

These proofs are simpler than the result for three dimensions, and relate to two incredibly dense packings of spheres, called the E8 lattice in 8-dimensions and the Leech lattice in 24 dimensions.

This is all quite magical, but is it useful? In the 1960s an engineer called Gordon Lang believed so. Lang was designing the systems for modems and was busy harvesting all the mathematics he could find.

He needed to send a signal over a noisy channel, such as a phone line. The natural way is to choose a collection of tones for signals. But the sound received may not be the same as the one sent. To solve this, he described the sounds by a list of numbers. It was then simple to find which of the signals that might have been sent was closest to the signal received. The signals can then be considered as spheres, with wiggle room for noise. To maximize the information that can be sent, these 'spheres' must be packed as tightly as possible.

In the 1970s, Lang developed a modem with 8-dimensional signals, using E8 packing. This helped to open up the Internet, as data could be sent over the phone, instead of relying on specifically designed cables. Not everyone was thrilled. Donald Coxeter, who had helped Lang understand the mathematics, said he was "appalled that his beautiful theories had been sullied in this way".

JUAN PARRONDO &amp; NOEL-ANN BRADSHAW

## From paradox to pandemics

University of Madrid; University of Greenwich, London

In 1992, two physicists proposed a simple device to turn thermal fluctuations at the molecular level into directed motion: a 'Brownian ratchet'. It consists of a particle in a flashing asymmetric field. Switching the field on and off induces the directed motion, explained Armand Ajdari of the School of Industrial Physics and Chemistry in Paris and Jacques Prost of the Curie Institute in Paris.

Parrondo's paradox, discovered in 1996 by one of us (J.P.), captures the essence of this phenomenon mathematically, translating it into a simpler and broader language: gambling games. In the paradox, a gambler alternates between two games, both of which lead to an expected loss in the long term. Surprisingly, by switching between them, one can

produce a game in which the expected outcome is positive. The term 'Parrondo effect' is now used to refer to an outcome of two combined events being very different from the outcomes of the individual events.

A number of applications of the Parrondo effect are now being investigated in which chaotic dynamics can combine to yield non-chaotic behaviour. For example, the effect can be used to model the population dynamics in outbreaks of viral diseases and offers prospects of reducing the risks of share-price volatility. Plus it plays a leading part in the plot of Richard Armstrong's 2006 novel, *God Doesn't Shoot Craps: A Divine Comedy*.

PETER ROWLETT

## From gamblers to actuaries

University of Birmingham, UK

In the sixteenth century, Girolamo Cardano was a mathematician and a compulsive gambler. Tragically for him, he squandered most of the money he inherited and earned. Fortunately for modern actuarial science, he wrote in the mid-1500s what is considered to be the first work in modern probability theory, *Liber de ludo aleae*, finally published in a collection in 1663.

Around a century after the creation of this theory, another gambler, Chevalier de Méré, had a dilemma. He had been offering a game in which he bet he could throw a six in four rolls of a die, and had done well out of it. He varied the game in a way that seemed sensible, betting he could throw a double six with two dice in 24 rolls. He had calculated the chances of winning in both games as equivalent, but found he lost money in the long run playing the second game. Confused, he asked his friend Blaise Pascal for an explanation. Pascal wrote to Pierre de Fermat in 1654. The ensuing correspondence laid the foundations for probability theory, and when Christiaan Huygens learned of the results he wrote the first published work on probability, *De Ratiociniis in Ludo Aleae* (published 1657).

In the late seventeenth century, Jakob Bernoulli recognized that probability theory could be applied much more widely than to games of chance. He wrote *Ars Conjectandi* (published, after his death, in 1713), which consolidated and extended the probability work by Cardano, Fermat, Pascal and Huygens. Bernoulli built on Cardano's discovery that with sufficient rolls of a fair, six-sided die we can expect each outcome to appear around one-sixth of the time, but that if we roll one die six times we shouldn't expect to see each outcome precisely once. Bernoulli gave a proof of the law of large numbers,

which says that the larger a sample, the more closely the sample characteristics match those of the parent population.

Insurance companies had been limiting the number of policies they sold. As policies are based on probabilities, each policy sold seemed to incur an additional risk, the cumulative effect of which, it was feared, could ruin a company. Beginning in the eighteenth century, companies began their current practice of selling as many policies as possible, because, as Bernoulli's law of large numbers showed, the bigger the volume, the more likely their predictions are to be accurate.

## JULIA COLLINS

### From bridges to DNA

*University of Edinburgh, UK*

When Leonhard Euler proved to the people of Königsberg in 1735 that they could not traverse all of their seven bridges in one trip, he invented a new kind of mathematics: one in which distances didn't matter. His solution relied only on knowing the relative arrangements of the bridges, not on how long they were or how big the land masses were. In 1847, Johann Benedict Listing finally coined the term 'topology' to describe this new field, and for the next 150 years or so, mathematicians worked to understand the implications of its axioms.

For most of that time, topology was pursued as an intellectual challenge, with no expectation of it being useful. After all, in real life, shape and measurement are important: a doughnut is not the same as a coffee cup. Who would ever care about 5-dimensional holes in abstract 11-dimensional spaces, or whether surfaces had one or two sides? Even practical-sounding parts of topology such as knot theory, which had its origins in attempts to understand the structure of atoms, were thought to be useless for most of the nineteenth and twentieth centuries.

Suddenly, in the 1990s, applications of topology started to appear. Slowly at first, but gaining momentum until now it seems as if there are few areas in which topology is not used. Biologists learn knot theory to understand DNA. Computer scientists are using braids — intertwined strands of material running in the same direction — to build quantum computers, while colleagues down the corridor use the same theory to get robots moving. Engineers use one-sided Möbius strips to make more efficient conveyor belts. Doctors depend on homology theory to do brain scans, and cosmologists

use it to understand how galaxies form. Mobile-phone companies use topology to identify the holes in network coverage; the phones themselves use topology to analyse the photos they take.

It is precisely because topology is free of distance measurements that it is so powerful. The same theorems apply to any knotted DNA, regardless of how long it is or what animal it comes from. We don't need different brain scanners for people with different-sized brains. When Global Positioning System data about mobile phones are unreliable, topology can still guarantee that those phones will receive a signal. Quantum computing won't work unless we can build a robust system impervious to noise, so braids are perfect for storing information because they don't change if you wiggle them. Where will topology turn up next?



## CHRIS LINTON

### From strings to nuclear power

*Loughborough University, UK*

Series of sine and cosine functions were used by Leonard Euler and others in the eighteenth century to solve problems, notably in the study of vibrating strings and in celestial mechanics. But it was Joseph Fourier, at the beginning of the nineteenth century, who recognized the great practical utility of these series in heat conduction and began to develop a general theory. Thereafter, the list of areas in which Fourier series were found to be

useful grew rapidly to include acoustics, optics and electric circuits. Nowadays, Fourier methods underpin large parts of science and engineering and many modern computational techniques.

However, the mathematics of the early nineteenth century was inadequate for the development of Fourier's ideas, and the resolution of the numerous problems that arose challenged many of the great minds of the time. This in turn led to new mathematics. For example, in the 1830s, Gustav Lejeune Dirichlet gave the first clear and useful definition of a function, and Bernhard Riemann in the 1850s and Henri Lebesgue in the 1900s created rigorous theories of integration. What it means for an infinite series to converge turned out to be a particularly slippery animal, but this was gradually tamed by theorists such as Augustin-Louis Cauchy and Karl Weierstrass, working in the 1820s and 1850s, respectively. In the 1870s, Georg Cantor's first steps towards an abstract theory of sets came about through analysing how two functions with the same Fourier series could differ.

The crowning achievement of this mathematical trajectory, formulated in the first decade of the twentieth century, is the concept of a Hilbert space. Named after the German mathematician David Hilbert, this is a set of elements that can be added and multiplied according to a precise set of rules, with special properties that allow many of the tricky questions posed by Fourier series to be answered. Here the power of mathematics lies in the level of abstraction and we seem to have left the real world behind.

Then in the 1920s, Hermann Weyl, Paul Dirac and John von Neumann recognized that this concept was the bedrock of quantum mechanics, since the possible states of a quantum system turn out to be elements of just such a Hilbert space. Arguably, quantum mechanics is the most successful scientific theory of all time. Without it, much of our modern technology — lasers, computers, flat-screen televisions, nuclear power — would not exist. ■

#### CORRECTIONS

In the Comment article 'Buried by bad decisions' (*Nature* **474**, 275–277), the statement "we will save lives by pushing a trolley into a person but not a person into a trolley" refers to an incorrect reference. The correct one is J. D. Greene *et al.* *Science* **293**, 2105–2108 (2001).

The Comment article 'Crowd control in Rwanda' (*Nature* **475**, 572–573) should have stated that family-planning aid dropped from 30% to 12% of overall health aid, not overall aid.





The splendid tail feathers of a peacock are displayed during courtship; feathers with the same basic structure fulfil roles from aerodynamics to insulation.

## EVOLUTION

# A tale of feathers

**Alan Brush** enjoys a compelling narrative on the discoveries that have illuminated the complexities and evolution of plumage.

**T**his book, about the natural history of feathers, begins with *Archaeopteryx*. This late-Jurassic (about 150-million-year-old) fossil, something between a reptile and bird, confounded and delighted evolutionists Charles Darwin and Thomas Huxley and palaeontologist Richard Owen. A small beast with reptilian teeth, long tail and skeletal features of both groups, it had clearly identifiable feathers with modern shape and structure. *Archaeopteryx* feathers were identical to those that today fascinate laymen, ornithologists, fashionistas and casual collectors.

Thor Hanson's storytelling is enhanced by his infectious excitement. In *Feathers*, he interviews the leading proponents on all sides of the controversies that surround the

origin and evolution of feathers and the birds that produce them.

Hanson explains the physics of how feather structures interact with light to produce amazing iridescent colours. He catalogues how different feathers with the same hollow, branching structure provide insulation (down), protection (contour), aerodynamic surfaces (wing and tail) and sensory input (filoplumes). Plumage helps with species identification, dictates behaviour and provides the spectacular decorations that birders enjoy. Hanson also traces the long arguments between advocates of the 'ground-up' and 'tree-down' theories of how the first birds took

to the air, and the alternative 'wing-assisted incline running' hypothesis.

Thanks to feathers' myriad qualities, people have used them as quill pens to sign significant historical papers, as badges and advertisements, on clothing and in fishing flies and as stuffing for garments, mattresses, quilts and cushions. Hanson's style makes the concepts of chemical morphogenesis, coherent scattering and reaction-diffusion waves accessible. His narrative is accompanied by a small number of diagrams and images and an appendix illustrating feather types. That it is not a picture book is an accomplishment when writing about colourful plumage and exotic behaviours.

In the 1970s, debate on the evolutionary origin of birds was reinvigorated by the

**NATURE.COM**  
Biologist Nicky Clayton on birds and the tango:  
[go.nature.com/tawyyyn](http://go.nature.com/tawyyyn)

P. D. STEWART/SPL





**Feathers: The Evolution of a Natural Miracle**

THOR HANSON  
Basic Books: 2011.  
352 pp. \$25.99

US palaeontologist John Ostrom, who posited that birds, as vertebrates with feathers, were related to theropod dinosaurs. Ostrom's claim was based on fossil evidence and supplemented by others' work on metabolism and behaviour. But for decades, the controversial argument that birds and dinosaurs were related

lacked a key element: an evolutionary history of feathers.

At the time, the dogma was that all birds — and only birds — have feathers. This changed in the 1990s, with the discovery of 'feathered dinosaurs' from the Yixian Formation in Liaoning Province in China. The rich fossil findings stimulated a re-evaluation of the evolutionary history of both feathers and the animals that bear them. Phylogenetic analysis confirmed that theropods and birds are sister groups, and the feather structures on the Yixian fossils provided direct evidence for the evolution of feathers. These findings complemented other data from ontogeny, molecular biology and morphology. Finally, a clear picture of the evolution of feathers has emerged.

Hanson's tale is comprehensive, accurate, timely and engaging. One thing missing is the story of the technical breakthroughs that led to the understanding of feather structure (keratin) and genomics. The fact that feathers are insoluble is partly because of their structure — they are made from highly organized filaments — and partly because of their amino-acid composition (they contain many stable intra- and intermolecular disulphide bonds).

In the late 1960s, a group in the protein-chemistry division of the Commonwealth Scientific and Industrial Research Organisation in Australia isolated and identified the soluble monomer of feather keratin, and revealed the characteristics of the gene family involved. Ornithologists quickly became interested. This accomplishment provided ways to test directly the 'feathers arose from scales' hypothesis and to map molecular evolution more widely onto lineages derived from other features. Comparative work on the proteins of the other epidermal structures, such as claws, scales and beaks, soon followed.

*Feathers* is a compelling introduction to one of nature's wonders. ■

**Alan Brush** is professor emeritus in ornithology at the University of Connecticut, Storrs, USA.  
e-mail: brushes2@juno.com

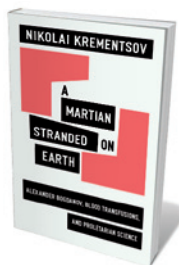
## Books in brief



**Sex on the Moon: The Amazing Story Behind the Most Audacious Heist in History**

Ben Mezrich DOUBLEDAY 320 pp. \$26.95 (2011)

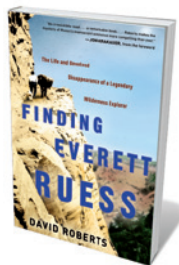
In 2002, NASA fellow Thad Roberts, aided by three interns, stole lunar and martian samples from a Johnson Space Center vault in Houston, Texas. As writer Ben Mezrich deftly recounts, Roberts's motivation was not geological obsession, but a desire to impress one of his accomplices, Tiffany Fowler. In a bizarre act that was both poetic and literal-minded, Roberts made love to her on a bed strewn with Moon rocks — hence the book's title. Rarely has career suicide been so entertaining.



**A Martian Stranded on Earth: Alexander Bogdanov, Blood Transfusions, and Proletarian Science**

Nikolai Kremontsov UNIVERSITY OF CHICAGO PRESS 192 pp. \$35 (2011)

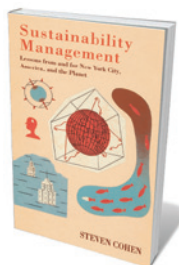
We sometimes forget that the Russian revolution convulsed science as well as society. Now philosopher of science Nikolai Kremontsov gives a portrait of a Bolshevik scientist at the epicentre of that revolution. A political rival to Lenin, Alexander Bogdanov was a physician, philosopher and sci-fi writer. Kremontsov sketches a rounded picture of a polymath who set up the world's first institute for blood-transfusion research and whose philosophical work laid the foundations of systems theory.



**Finding Everett Ruess: The Life and Unsolved Disappearance of a Legendary Wilderness Explorer**

David Roberts BROADWAY BOOKS 416 pp. \$24.99 (2011)

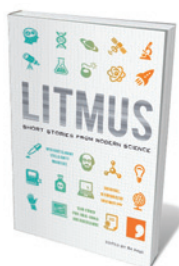
American wilderness artist and writer Everett Ruess, a contemporary of photographer Ansel Adams, was an archaeologist-naturalist manqué who ventured solo into remote areas of Arizona, Colorado, New Mexico and Utah, starting when he was just 15. Ruess was a prodigious journal-keeper, poet and printmaker, but disappeared in November 1934 near Escalante, Utah, aged just 20. His fate remains a mystery but his works continue to astound. Roberts shows that we can still 'find' Ruess in compilations of his art and writings.



**Sustainability Management: Lessons from and for New York City, America, and the Planet**

Steven Cohen COLUMBIA UNIVERSITY PRESS 256 pp. \$35 (2011)

Some 25 years after the concept of sustainability emerged, policy solutions to implementing it remain elusive. Cohen, executive director of Columbia University's Earth Institute in New York, argues that we now have enough successful examples to draw up blueprints for keeping the planet viable and economies afloat. Through case studies such as New York's community gardens, Cohen looks at sustainable practice in business, energy, water and food supply, and the technical, financial and political challenges of transmuting ideas into action.



**Litmus: Short Stories from Modern Science**

Edited by Ra Page COMM PRESS 298 pp. £9.99 (2011)

From Jeremiah Horrocks's observation of the transit of Venus in 1639 to Alan Turing's revelations about morphogenesis in 1952, 'eureka' moments shift our take on the cosmos. They can also make for supercharged narratives. In 17 short stories, novelists and poets including Sean O'Brien and Kate Clanchy retell lightbulb moments from centuries of science. Each has an afterword by an expert, from Jim Al-Khalili on Einstein's special theory of relativity to Denis Noble on heart modelling and Giacomo Rizzolatti on mirror neurons.



## PALAEOLOGY

# Living it large

**Brian Switek** swoons over a New York exhibition that brings giant sauropods back to life.

Dinosaur halls are the petrified trophy rooms of natural-history museums. But *The World's Largest Dinosaurs* — an exhibition curated by Mark Norell of the American Museum of Natural History and Martin Sander of Germany's University of Bonn — is not a classic gallery of old bones.

The exhibition reconstructs the lives of *Apatosaurus*, *Brachiosaurus* and other sauropods by showcasing recent research into their biology. Instead of ranks of enormous skeletons, visitors are greeted with the restored head and neck of *Argentinosaurus* (the full 30-metre body would be too big) and an 18-metre *Mamenchisaurus* mock-up, which stands in a pathway of interactive displays.

Walk along the left flank of the *Mamenchisaurus* and you will see its ribs expand and contract as it breathes. Linger a little longer and its flesh peels back to reveal and explain the creature's internal anatomy.

Casts and fragments

**NATURE.COM**  
For more on large  
vertebrate fossils:  
[go.nature.com/q49vxo](http://go.nature.com/q49vxo)

of other sauropods illustrate how these dinosaurs coped with being so large. How could giants such as *Diplodocus*, for example, eat enough food given their tiny skulls and small, peg-like teeth? The displays show that, in fact, their heads were well-suited to living large. Sauropods bolted down vast quantities of food without chewing, and their long necks allowed them to sample wide swathes of greenery while standing still.

For how long their meals would have filled them up is another matter. An interactive display invites visitors to select a warm-blooded (endothermic) or cold-blooded (ectothermic) sauropod and choose a diet of either high- or low-quality plants — such as cycads or horsetails, respectively. As you 'feed' the sauropod, the dinosaur's virtual stomach fills up at different rates and for varying amounts of time, illustrating how diet and physiology interact.

Just one part of the exhibit feels out of

**The World's  
Largest Dinosaurs**  
American Museum of  
Natural History,  
New York.  
Until 2 January 2012.

place. Following on from kiosks about blood pressure, growth rates and lung anatomy, the final room houses a trough filled with artificial dinosaur bones for children to excavate. Although it reminds visitors that fieldwork is the first step in understanding prehistoric life, it jars with the exhibition's focus.

Naturally, reverse-engineering the anatomy and physiology of animals from prehistoric bones involves speculation and informed guesswork. What sauropod hearts looked like must be inferred from those of birds and crocodiles, and the physiological functions of the air sacs in sauropod bones are still debated. Yet this exhibit is a fitting tribute to how palaeontology has matured.

A century ago, when institutions such as the American Museum of Natural History were new, palaeontologists competed to find the biggest and most complete sauropod skeletons to display. *The World's Largest Dinosaurs* shows how the study of these animals has become an interdisciplinary science that is beginning to answer long-standing questions about dinosaur biology. It is a wonderful celebration of the efforts of palaeontologists to put flesh on ancient bones. ■ **SEE NEWS FEATURE P.159**

**Brian Switek** is a freelance science writer and author of *Written in Stone: Evolution, the Fossil Record, and Our Place in Nature*. e-mail: [evogeek@gmail.com](mailto:evogeek@gmail.com)



The muscles and internal organs of an 18-metre *Mamenchisaurus* are on show at the American Museum of Natural History in New York.



Psychologist Herbert Terrace with Nim Chimpsky in the 1970s, and today (below).

## Q&A Herbert Terrace

# The interpreter

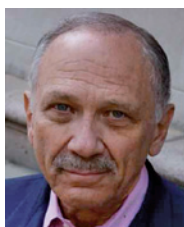
*In 1973, Herbert Terrace, a psychologist at Columbia University in New York, embarked on an experiment to teach sign language to an infant chimpanzee named Nim Chimpsky, after linguist Noam Chomsky. On the release of the documentary Project Nim, Terrace talks about research ethics, chimp cognition and the origins of language.*

### How did the experiment come about?

After serving as a graduate assistant at Harvard University with behavioural psychologist B. F. Skinner, I heard that Allen and Beatrix Gardner at the University of Nevada, Reno, were teaching sign language to a chimpanzee named Washoe. But when I looked at their data, I wasn't sure that the chimp's sequences of hand signs were grammatical. I decided to do a study to collect everything a chimp signed, and document the circumstances. We wanted to have full records of the discourse between the infant chimp and the caretaker.

### How did you get started?

I went to the Institute for Primate Studies in Norman, Oklahoma, and the director offered me a newborn chimp. My PhD student volunteered the use of her New York townhouse, and we tried to immerse the young chimp in a sign-language-only environment, although neither of us was fluent. Then the president of Columbia University provided a mansion in Riverdale in exchange for us paying the heating



bill. The project shifted into the hands of Laura Ann Petitto, an enthusiastic student who kept good records and who is now a cognitive neuroscientist studying language at Gallaudet University in Washington DC.

### Did the experiment meet your expectations?

The language didn't materialize. A human baby starts out mostly imitating, then begins to string words together. Nim didn't learn. His three-sign combinations — such as 'eat me eat' or 'play me Nim' — were redundant. He imitated signs to get rewards. I published the negative results in 1979 in the journal *Science*, which had a chilling effect on the field.

### Why couldn't Nim put a sentence together?

I haven't seen any evidence that a chimp has a theory of mind. It can predict behaviour, but the concept of another individual's thinking is foreign to it. So it is pointless for a chimp to start a conversation: why talk unless you expect a reply? Rhesus macaques are able to learn a long sequence of images by trial and error, but no one has accused them of having language. Even if you could get a chimp to learn calculus, it will never converse.

**Project Nim**  
DIRECTED BY JAMES MARSH  
In US cinemas now.  
UK release on 12 August.

Humans have language not because we are smart, but because we are social and sensitive to the thoughts of others.

### How did the experiment end?

Nim was getting bigger, and you couldn't take the chimp out of him. He knew when people were afraid, and he would bite them if they weren't confident. So I flew him back to Oklahoma. Most people were angry at me when I called off the project. They thought it was cruel. And they were right: it is emotionally wrenching to socialize a chimp as a child and then put him back with chimps. But what is the alternative? If I were to undertake this again, I would give more thought to what happens when the project ends. You have to find a permanent home for him, a facility with space for a few chimps, and visits from the original caretakers. It is unkind if you don't have an exit strategy.

### What became of Nim?

When the institute in Oklahoma ran out of funding, they sold off their chimps for medical research. But I felt that to have an intelligent and well-trained chimp subjected to hepatitis research was immoral. The president of New York University, which operated the medical-research facility, set Nim free. He lived the rest of his life in a giant cage at a ranch for celebrity animals in Texas. In 2000, at the age of 26, he died of a heart attack.

### Did Nim's fate change your views on the ethics of animal research?

Some people want to abolish animal research, but I don't think they understand the scientific loss and the implications for the welfare of humans. If you can get the same information without using an animal, you shouldn't use one. But there are medical experiments in which you have to sacrifice an animal. These must be done as humanely as possible. Other experiments can enhance the animal's environment. I've taught monkeys video games to test their memory. After a vacation they're raring to go again. To get good data, you want the animal to be happy.

### What did you think of the film *Project Nim*?

I thought it was brilliant but I had some misgivings. There wasn't a concise explanation of why you would want to teach sign language to a chimp, or why Nim couldn't acquire language. It portrayed me as an absentee landlord and suggests that I didn't care emotionally about him. This was not true: I drove him to his mansion and spent time playing and signing with him. But he was the subject of a scientific study, and I emphasized the scientific goal. I did not think of him as a child. ■

INTERVIEW BY JASCHA HOFFMAN



# CORRESPONDENCE

## Time to end US chimp studies

As a physician, medical educator and former animal researcher, I agree that the US national discussion of chimpanzee experimentation should go beyond simple husbandry issues (*Nature* 474, 252; 2011). For ethical and scientific reasons, it is time for the United States to join other developed countries in ending invasive experiments on chimpanzees.

We now know a great deal about the awareness, intelligence and emotional responses of chimpanzees. Scientists from my organization, for example, have found that invasive experiments on chimpanzees can induce symptoms of depression, anxiety and compulsive behaviours that are similar to mood and anxiety disorders seen in traumatized humans (H. R. Ferdowsian *et al.* *PLoS ONE* 6, e19855; 2011).

Although chimps are humankind's closest genetic relatives, we show significant differences in our gene expression, physiology and disease susceptibility. It is becoming increasingly evident that chimpanzee experiments are not improving our understanding and treatment of human disease. Billions of dollars and decades of research using chimpanzees have not produced effective vaccines for hepatitis C, HIV, malaria or other diseases, nor have they provided insight into cancer, neurological diseases or psychiatric disorders. However, the process has inflicted extensive and often lifelong pain and suffering on these animals.

**John J. Pippin** *Physicians Committee for Responsible Medicine, Washington DC, USA.*  
[jpippin@pcrm.org](mailto:jpippin@pcrm.org)

## *E. coli* outbreak exposed tech gaps

As well as the organizational mismanagement of the recent *Escherichia coli* outbreak in Germany (*Nature* 474, 251; 2011), the technical underdevelopment of the country's medical microbiology institutes is staggering, given that Germany is the largest economy in Europe. Such shortcomings leave the country unprotected against attacks by highly virulent agents of natural or bioterrorist origin.

For example, none of these institutes is set up for rapid sequencing or mass screening of major pathogenic agents using the polymerase chain reaction (PCR). Most of the labs still rely solely on Robert Koch's lengthy culture methods, even though analysis of a known pathogen could be reduced to a few hours by using culture enrichment combined with high-throughput real-time PCR. Such an analysis during the recent outbreak would have increased the number of samples tested and probably saved lives.

**Rainer Fislage** *St Wendel, Germany.*  
[rainer.fislage@biophenium.de](mailto:rainer.fislage@biophenium.de)

## Dam not sole cause of Chinese drought

China's Yangtze River is suffering its worst drought for more than 50 years (*Nature* doi:10.1038/news.2011.315; 2011). Although people blame the Three Gorges Dam for making matters worse, other factors have also contributed.

In response to April and May's severe water shortage, central government ordered the release of 50 billion cubic metres of water from reservoirs in the Yangtze basin, most of

which came from the Three Gorges Reservoir, with some from other reservoirs used for power generation. The result is a cumulative toll on the river's large-scale hydrological balance.

Extensive land reclamation in the middle and lower reaches of the river has exacerbated the drought by removing or shrinking many natural lakes across the river basin. Worse, more than 80% of the remaining lakes are no longer connected with the river, seriously limiting their capacity for buffering the water supply.

Excessive pumping of groundwater is a significant contributor to the current drought (see [go.nature.com/b6Srkg](http://go.nature.com/b6Srkg); in Chinese), as are channel incisions caused by loss of sediment and sand mining.

The drought's severity threatens China's south-north water-diversion project, a huge trans-basin scheme to ease the water shortage in northern China.  
**X. X. Lu, Xiankun Yang, Siyue Li** *National University of Singapore, Singapore.*  
[geoluxx@nus.edu.sg](mailto:geoluxx@nus.edu.sg)

## Green labelling being misused

'Green' labelling of items produced sustainably has become a much-publicized tool of the environmental movement. But green-label criteria that were developed for forestry are now being inappropriately applied to agricultural crops — with unacceptable risks to wildlife.

The high conservation value (HCV) concept was originally intended for timber harvested without harming large forest blocks or critically endangered flora and fauna. It has since been extended to some of the most rapidly expanding crops in the tropics — namely, oil palm, soya bean, sugar cane and cacao — ostensibly to ensure their

'sustainable' production.

The combined area of these four crops increased by 36.5 million hectares between 1999 and 2008, largely in countries of exceptional biodiversity, such as Brazil and Indonesia, and largely at the expense of forest. The hope is that green labelling through HCV will stem this tide of habitat destruction and biodiversity loss.

But many of the world's most intact and biodiverse tropical forests (including the Amazon Basin, Congo and New Guinea) harbour few critically endangered species. In such cases, the only HCV criterion that is likely to prevent forest conversion to agriculture is the one protecting large expanses of habitat.

Unfortunately, the round tables for oil palm, soya bean, sugar cane and cacao have decided that only forest blocks larger than 20,000–500,000 hectares (depending on the country) are eligible for protection under this criterion. So crops replacing areas of forest below these thresholds can perversely carry a green label, even though these thresholds are dangerously high for wildlife survival.

The size of HCV thresholds must be drastically reduced so that green-labelled crops are not simply 'greenwash'.

**David P. Edwards, Brendan Fisher, David S. Wilcove** *Princeton University, New Jersey, USA.*  
[dpedward@princeton.edu](mailto:dpedward@princeton.edu)

### CONTRIBUTIONS

Correspondence may be sent to [correspondence@nature.com](mailto:correspondence@nature.com) after consulting the author guidelines at [go.nature.com/cmchno](http://go.nature.com/cmchno). Readers are also welcome to comment online on anything published in *Nature*: [www.nature.com/nature](http://www.nature.com/nature).

## Weight loss through smoking

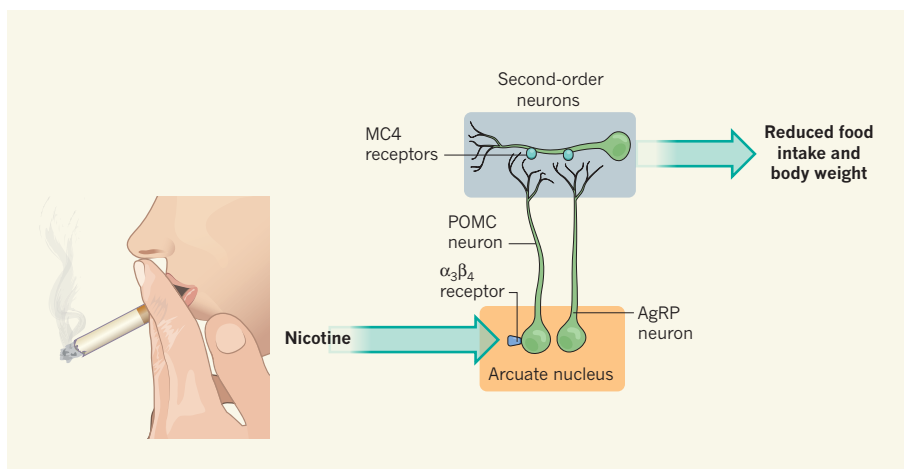
Many people smoke to keep their weight down. The identification of the molecular target in the brain for the appetite-suppressant effects of nicotine is a first step towards finding healthy alternatives to smoking for weight management.

RANDY J. SEELEY & DARLEEN A. SANDOVAL

Anyone who has tried to stop smoking knows that quitting is frequently followed by a rapid weight gain<sup>1</sup> of some 4 or 5 kilograms, and as much as 13 kilograms. It isn't surprising, therefore, that despite a general decrease in the rates of smoking within the past decade<sup>2</sup>, it remains prevalent in professions such as acting and modelling, in which being thin is often a job requirement. Indeed, the prospect of a gain in weight is a major deterrent to many smokers who want to quit but who see it either as a threat to their appearance or, wrongly<sup>3</sup>, as a greater threat to their health than smoking. Writing in *Science*, Mineur *et al.*<sup>4</sup> shed light on the molecular targets of nicotine in the brain that are responsible for the compound's ability to suppress appetite. Their data may lead to alternative — healthy — ways of maintaining reduced body weight.

Over the past 15 years, our understanding of the brain circuitry that serves to control appetite and regulate body weight has grown rapidly. One of the most notable advances has been the discovery that the melanocortin (MC) system in the brain seems to be a crucial regulator of body weight<sup>5</sup>. In peripheral tissues of the body, this system regulates such features as the colour of hair and skin. However, genetic analyses in humans and animals, as well as pharmacological data, have established that activation of the MC4 receptor in the brain reduces food intake and promotes weight loss<sup>5</sup>. Manipulations that reduce the activity of this receptor are associated with increased food intake and weight gain<sup>5</sup>.

MC4-receptor activity is regulated by a complex interplay between two populations of neurons; these cells are found intermingled in a small brain region called the arcuate nucleus within the hypothalamus. One neuron population (POMC) synthesizes the precursor to several molecules that activate MC4 receptors. The other population (AgRP) synthesizes an endogenous antagonist to these receptors (Fig. 1). When individuals lose weight by restricting their calorie intake, decreased activity of the POMC neurons and increased activity of the AgRP neurons result in reduced



**Figure 1 | Appetite-suppressant effects of nicotine in the brain.** Mineur *et al.*<sup>4</sup> show that inhaled nicotine affects  $\alpha_3\beta_4$  nicotinic acetylcholine receptors located on POMC neurons in the arcuate nucleus of the hypothalamus. POMC neurons make a precursor of several agonist molecules that activate MC4 receptors, whereas AgRP neurons located in the same brain region make an antagonist molecule of such receptors. These endogenous molecules regulate the activity of MC4 receptors on second-order neurons involved in regulating food intake, powerfully suppressing appetite and so weight gain.

activity of MC4 receptors, which predisposes the individuals to regain the lost weight<sup>5</sup>.

Mineur *et al.*<sup>4</sup> explored the hypothesis that nicotine directly activates POMC neurons. They found that, indeed, receptors known as  $\alpha_3\beta_4$  nicotinic acetylcholine receptors, located on these neurons, mediate nicotine's potent appetite-suppressant effects. Specifically, nicotine-induced activation of these receptors enhances the firing of POMC neurons. The authors also show that in mice lacking these neurons, nicotine no longer suppresses appetite.

Mineur and colleagues' data are supported by another study<sup>6</sup> which also shows that not only does nicotine lead to increased activity of POMC neurons, but it also enhances their synaptic communication with other neurons. Together, these results strongly support the idea that nicotine exerts its potent effects via the brain MC system, and that the weight gain associated with cessation of smoking is due to reduced activity of MC4 receptors.

These findings<sup>4</sup> have two major implications. First, they further confirm the central role of the brain MC system in weight regulation through MC4 receptors. In fact, these data put nicotine in good company with several other successful ways of manipulating

letters of the alphabet appetite that seem to exert their effects via the MC system. The effects on food intake of many things — from bariatric surgery to a high-fat diet to endogenous peripheral hormones such as leptin, ghrelin and PYY — are thought to be mediated by changes in this brain system.

In addition, the results indicate a potentially new way of activating the MC system to achieve weight loss. Obesity remains the largest unmet medical challenge in developed countries, and yet no new drug has been approved in the United States to treat this disorder since 1999. Obviously, smoking is far from an ideal way to avoid obesity, but targeting specific receptors on POMC neurons that mediate the appetite-suppressing effect of nicotine may be a pharmacological approach to providing obese individuals with safe and sustained weight loss. This is particularly important because, although drugs that target MC4 receptors directly are effective weight-loss agents, they are plagued by side effects, including increased heart rate and blood pressure<sup>7</sup>. Stimulating the release of endogenous activators of MC4 would provide an attractive alternative by targeting more discrete populations of these receptors. ■



Randy J. Seeley and Darleen A. Sandoval are at the Metabolic Diseases Institute, University of Cincinnati, Cincinnati, Ohio 45237, USA. e-mail: randy.seeley@uc.edu

1. Chiolerio, A., Faeh, D., Paccaud, F. & Cornuz, J. *Am. J. Clin. Nutr.* **87**, 801–809 (2008).
2. www.cdc.gov/mmwr/preview/mmwrhtml/mm5935a3.htm

3. Williamson, D. F. et al. *N. Engl. J. Med.* **324**, 739–745 (1991).
4. Mineur, Y. S. et al. *Science* **332**, 1330–1332 (2011).
5. Schwartz, M. W. et al. *Nature* **404**, 661–671 (2000).
6. Huang, H., Xu, Y. & van den Pol, A. J. *Neurophysiol.* doi:10.1152/jn.00740.2010 (2011).
7. Greenfield, J. R. et al. *N. Engl. J. Med.* **360**, 44–52 (2009).

## GLOBAL CHANGE

# Indirect feedbacks to rising CO<sub>2</sub>

There have been many studies on the effects of enriched levels of atmospheric carbon dioxide on soils. A meta-analysis shows that emissions of other greenhouse gases increase under high-CO<sub>2</sub> conditions. [SEE LETTER P.214](#)

ALEXANDER KNOHL & EDZO VELDKAMP

Human activities have caused atmospheric concentrations of carbon dioxide, a major greenhouse gas, to increase at an accelerating pace. Starting at around 280 parts per million (p.p.m.) in pre-industrial times, they have now exceeded 390 p.p.m., and are expected to reach 600–800 p.p.m. by the end of the century<sup>1</sup>. On page 214 of this issue<sup>2</sup>, van Groenigen and colleagues add to our awareness of the complex consequences of this trend, in terms of the effect that it will have on emissions of other greenhouse gases from various ecosystems.

In producing global warming, CO<sub>2</sub> is responsible for the largest part of the anthropogenic impact on Earth's energy balance. It is, of course, also an essential nutrient for plant metabolism. Numerous CO<sub>2</sub>-enrichment experiments over the past two decades have demonstrated the positive effect of elevated CO<sub>2</sub> on plant growth — increased biomass and increased carbon storage in soils<sup>3</sup>. The vegetation response to elevated CO<sub>2</sub> might be constrained by various interactions with water and nutrients such as nitrogen<sup>4,5</sup>. However, experiments and model projections suggest that accelerated plant growth due to CO<sub>2</sub> fertilization could draw down some of this gas from the atmosphere, and hence could weaken future rates of CO<sub>2</sub> increase and lessen the severity of climate change<sup>6</sup>.

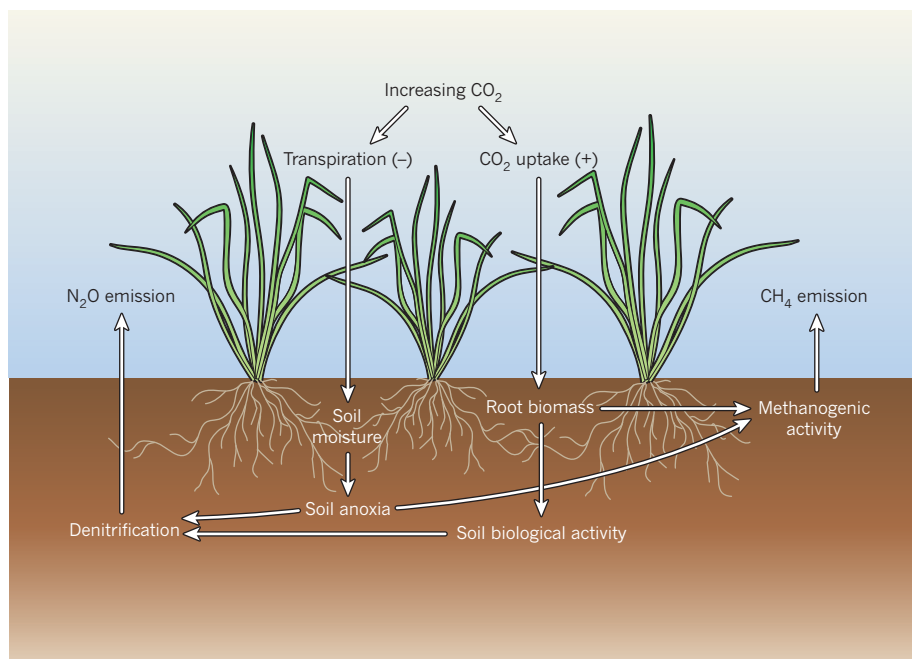
Van Groenigen *et al.*<sup>2</sup> present evidence that rising levels of CO<sub>2</sub> are not only resulting in an increased carbon sink in terrestrial ecosystems, but could also cause increased emissions of other, much more potent, greenhouse gases such as methane (CH<sub>4</sub>) and nitrous oxide (N<sub>2</sub>O) from soils. Methane is produced by anaerobic methanogenic microorganisms that thrive in wetlands, including rice paddies, where labile (biologically accessible) carbon is

available and diffusion of oxygen into the soil is severely restricted. Nitrous oxide is mainly produced in soils by aerobic nitrifying and anaerobic denitrifying bacteria. The interaction between nitrogen availability and soil water content controls the rate of N<sub>2</sub>O production. The respective global-warming potentials of CH<sub>4</sub> and N<sub>2</sub>O are 25 and 298 times greater than that of CO<sub>2</sub>, and thus they influence

Earth's energy balance even though they occur in much smaller concentrations.

Van Groenigen and colleagues collected information from 49 published studies that reported the effect of atmospheric CO<sub>2</sub> enrichment on CH<sub>4</sub> and N<sub>2</sub>O fluxes from soils. Using a meta-analysis, they show that elevated CO<sub>2</sub> stimulated N<sub>2</sub>O emissions by 18.8%, and that CH<sub>4</sub> emissions from wetlands increased by 13.2% and from rice paddies by as much as 43.4%. Notably, they also suggest the mechanisms that are probably responsible for these observed increases in greenhouse-gas emissions (Fig. 1).

Their suggestion goes as follows. Elevated CO<sub>2</sub> led to reduced plant transpiration (the evaporation of water from plant surfaces, leaves in particular), which increased soil water content and promoted the existence of anaerobic microsites in soils. This, together with increasing biological activity, probably stimulated denitrification and consequently N<sub>2</sub>O production. Also, the CO<sub>2</sub>-induced increase in root biomass may have contributed by increasing the availability of labile carbon, a crucial energy source for denitrification. The CO<sub>2</sub>-induced stimulation of CH<sub>4</sub> emissions from wetlands and rice paddies was probably the result of higher net plant production, leading to increasing carbon availability for substrate-limited methanogenic microorganisms. Extrapolating their



**Figure 1 | Proposed mechanisms of increased N<sub>2</sub>O and CH<sub>4</sub> emissions from soils.** From their meta-analysis, van Groenigen *et al.*<sup>2</sup> estimate that rising levels of atmospheric CO<sub>2</sub> will result in more output of N<sub>2</sub>O from upland soil (at a rate equivalent to 0.57 Pg CO<sub>2</sub> yr<sup>-1</sup>) and of CH<sub>4</sub> from rice paddies and wetlands (equivalent to 0.56 Pg CO<sub>2</sub> yr<sup>-1</sup>). They suggest that these increases are caused by reduced plant transpiration under conditions of elevated CO<sub>2</sub>, resulting in increased soil moisture. Together with increased root biomass, this leads both to greater denitrification (and hence increased N<sub>2</sub>O emission) and to more methanogenic activity (and hence increased CH<sub>4</sub> emission). The increase in these greenhouse gases will thus partially offset the predicted enhanced uptake of carbon by terrestrial ecosystems in a high-CO<sub>2</sub> world.

results to the global scale, van Groenigen *et al.*<sup>2</sup> estimate that the combined effect of stimulated N<sub>2</sub>O and CH<sub>4</sub> emissions could be equivalent to at least 1.12 Pg CO<sub>2</sub> yr<sup>-1</sup> (Pg = petagrams = 10<sup>15</sup> grams). This is around 17% of the expected increase of the terrestrial CO<sub>2</sub> sink as a result of higher CO<sub>2</sub> concentrations.

Earlier studies have shown that long-term carbon sequestration in a CO<sub>2</sub>-enriched atmosphere can be constrained by nitrogen availability<sup>5,6</sup>. Critics may wonder how these studies and van Groenigen and colleagues' analysis fit together, as it seems unlikely that denitrification would be stimulated by elevated CO<sub>2</sub> in nitrogen-limited ecosystems. This apparent discrepancy may be explained by the geographical bias in the present paper. The large majority of the 49 studies included in the meta-analysis were located in temperate regions, in areas — the United States, Europe, China and Japan — that are nowadays subject to considerable deposition of atmospheric nitrogen<sup>7</sup>. Some ecosystems included in the meta-analysis, such as agricultural areas receiving little or no fertilizer, and regions of natural vegetation, may thus have been subject to the input of considerable anthropogenic nitrogen through the atmosphere. Because nitrogen deposition is predicted to increase in the coming decades, the studies may therefore be more representative of future conditions, when nitrogen deposition will have become a global feature.

Another striking point is the almost complete lack of studies in the tropics and subtropics, where the strongest increases in nitrogen deposition are expected to occur<sup>7</sup>. Some tropical ecosystems may react differently from temperate ecosystems to elevated CO<sub>2</sub> concentrations. Many intact tropical forests tend to cycle large quantities of nitrogen<sup>8</sup>, and an increase in soil-moisture content may have strong effects on N<sub>2</sub>O emissions even without nitrogen deposition. Tropical grasslands are dominated by grasses using the C<sub>4</sub> photosynthetic pathway, which may improve their water-use efficiency to different extents from that of plants using the C<sub>3</sub> pathway. There is a clear need for field studies in these ecosystems, in order to improve our ability to evaluate the overall effect of elevated CO<sub>2</sub> on the budgets of greenhouse-gas emissions.

Obviously, the report by van Groenigen *et al.*<sup>2</sup> is not the end of the story, and future research may provide evidence of other feedbacks that have not yet been quantified or even hypothesized. Nevertheless, this study provides the first comprehensive analysis of available data that shows the importance of indirect feedbacks of elevated CO<sub>2</sub> on CH<sub>4</sub> and N<sub>2</sub>O emissions on a global scale. It is now up to the scientific community to include these feedbacks in global climate models and to fill in the large gaps in information that still exist. ■

**Alexander Knohl** is in the Section of Bioclimatology, and **Edzo Veldkamp** is in the Section of Soil Science of Tropical and Subtropical Ecosystems, University of Göttingen, 37077 Göttingen, Germany. e-mails: [aknohl@uni-goettingen.de](mailto:aknohl@uni-goettingen.de); [eveldeka@uni-goettingen.de](mailto:eveldeka@uni-goettingen.de)

1. Solomon, S. D. *et al.* (eds) *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2007).

2. van Groenigen, K. J., Osenberg, C. W. & Hungate, B. A. *Nature* **475**, 214–216 (2011).
3. Hungate, B. A. *et al.* *Glob. Change Biol.* **15**, 2020–2034 (2009).
4. Reich, P. B. *et al.* *Nature* **440**, 922–925 (2006).
5. Oren, R. *et al.* *Nature* **411**, 469–472 (2001).
6. Thornton, P. E., Lamarque, J.-F., Rosenbloom, N. A. & Mahowald, N. M. *Glob. Biogeochem. Cycles* **21**, GB4018, doi:10.1029/2006GB002868 (2007).
7. Galloway, J. N. *et al.* *Science* **320**, 889–892 (2008).
8. Hedin, L. O., Brookshire, E. N. J., Menge, D. N. L. & Barron, A. R. *Annu. Rev. Ecol. Evol. Syst.* **40**, 613–635 (2009).

## REGENERATIVE MEDICINE

# Drawing breath after spinal injury

**New work on a rat model suggests that, after spinal-cord injury, restoration of sustained and robust respiratory function is possible using strategies that promote both neuronal plasticity and regeneration. [SEE ARTICLE P.196](#)**

KATHERINE ZUKOR & ZHIGANG HE

**A**bove C4, breathe no more.' This is the memory aid that reminds medical students that damage to the spinal cord above the fourth cervical vertebra (C4) — that is, the neck — can interrupt breathing. Injuries at the cervical level are the most common type of spinal-cord injury and account for more than half of all cases. Individuals who survive such injuries usually need ventilators to breathe, and so face a host of complications to their overall health and quality of life. A study by Alilain *et al.*<sup>1</sup> on page 196 of this issue offers hope that we may one day know how to treat this problem, so that patients with spinal-cord injuries above C4 can breathe on their own.

Breathing rate, rhythm and depth are controlled automatically by specialized regions of the brainstem<sup>2</sup> (Fig. 1a). The neurons in these regions send their axonal processes down the spinal cord to control the activity of other neurons in the phrenic motor nuclei (PMN) of the cervical spinal cord (C3–C6). The axons of the PMN neurons form the phrenic nerves, which, in turn, innervate the muscles of the diaphragm. Thus, contraction and relaxation of the diaphragm enable rhythmic breathing. When the spinal cord is injured above the C4 level, axons connecting the brainstem to the PMN are damaged, and breathing is disrupted. To make matters worse, axons in the adult spinal cord do not regenerate well, one of the main reasons being the inhibitory environment of the injured spinal cord<sup>3</sup>.

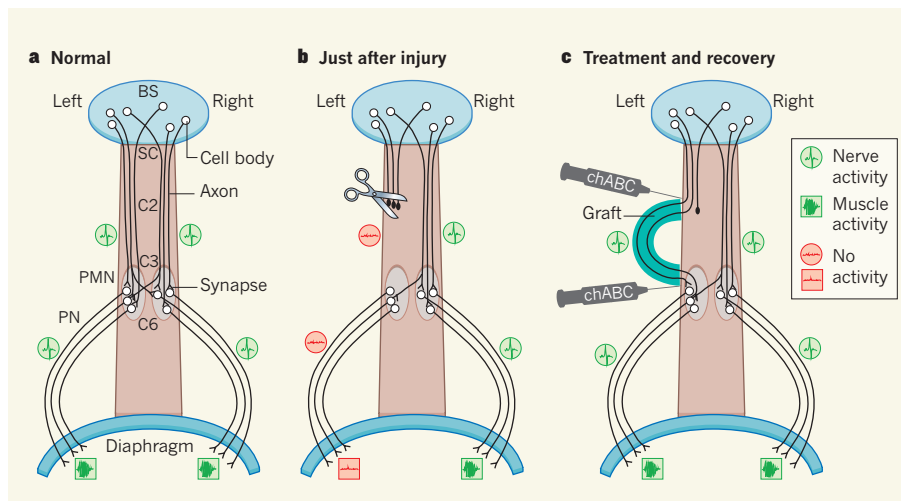
Over the years, researchers have invoked many strategies to provide axons with a more supportive environment. These include

either removing inhibitory molecules, such as chondroitin sulphate proteoglycans (CSPGs) in the extracellular matrix<sup>4</sup>, or grafting in a piece of peripheral nerve that could serve as a bridge for axonal growth<sup>5</sup>. Combinations of these approaches have yielded encouraging results. For example, after a cervical spinal-cord injury (SCI) in rats, applying a peripheral nerve graft, together with injection of the enzyme chondroitinase ABC (chABC) to degrade CSPGs, allows spinal-cord axons to regenerate through the graft, re-enter the spinal cord and form synaptic connections with neurons on the opposite side of the injury<sup>6</sup>.

Alilain *et al.*<sup>1</sup> applied a similar treatment strategy to recover respiratory function in rats after SCI. The authors made a partial injury at the C2 level to paralyse the diaphragm on one side of the animals' body (Fig. 1b). They then removed a piece of the rats' tibial nerve and grafted one end of it in the injury site at C2 and the other end in a small slit at the C4 level — near the PMN. Finally, they injected chABC at both ends of the graft, as well as in the PMN area, to degrade CSPGs (Fig. 1b).

Twelve weeks after injury, the group receiving this treatment had the highest percentage of recovered animals and the best quality of recovery in respiratory function compared with controls. Specifically, in many animals the paralysed half of the diaphragm muscle recovered nearly normal rhythmic electrical activity. Moreover, neurons from breathing centres of the brainstem grew axons into the graft. To demonstrate that recovery was largely due to axons that had regenerated through the graft and not just the rewiring of circuits in the portions of the spinal cord that were uninjured, Alilain *et al.* cut the graft; this treatment





**Figure 1 | Experimental manipulation of the respiratory system.** **a**, Basic anatomy of the intact respiratory system. BS, brainstem; SC, spinal cord; C, cervical level; PMN, phrenic motor nucleus; PN, phrenic nerve. **b**, After a partial spinal-cord injury at C2, innervation to the left PMN is interrupted and the diaphragm is paralysed on the left. **c**, Alilain *et al.*<sup>1</sup> show that 12 weeks after peripheral-nerve grafting and injection of chondroitinase ABC (chABC), activity is restored in the paralysed side of the diaphragm.

abolished the regained respiratory function.

Intriguingly, when the graft was cut, the residual electrical activity of the paralysed half of the diaphragm muscle was significantly, albeit transiently, increased compared with residual activity after the initial SCI. This suggests that spinal-cord circuits were also rewired to some extent in this experimental setting: descending regenerating axons may have connected with different targets, and denervated neurons may also have found new synaptic partners before regenerating axons could reach them<sup>7</sup>. It is encouraging that, despite the havoc such rewiring could wreak, the system could still adapt to the changes — perhaps through a type of learning process — to restore proper firing patterns to the motor neurons innervating the diaphragm.

Such adaptability might be a special property of the respiratory circuit: it has long been known<sup>8</sup> that there are latent connections in the respiratory system that can be activated after injury. Alternatively, it may be a common property of all spinal-cord circuits that can be unleashed by the degradation of CSPGs, as well as by extensive rehabilitation. At an axon's target site, such as the PMN, CSPGs are thought to stabilize circuits after development is complete<sup>9</sup>. Degradation of CSPGs after injury may therefore give circuits the flexibility they need to make new connections and adapt to changes. The built-in rehabilitation regimen that the paralysed diaphragm receives — by virtue of the fact that the animal must continue to breathe if it is to remain alive — probably also plays a part in shaping the new circuit. In support of this, chABC treatment together with physical rehabilitation promotes recovery of manual dexterity after cervical SCI in rats<sup>10</sup>.

Overall, Alilain and colleagues' results<sup>1</sup>

suggest that combinatorial strategies that promote both long-distance axon regeneration and local circuit reorganization may be universally useful for enabling functional recovery after SCI. Because not all axons of the central nervous system have the same ability to regrow into permissive grafts<sup>11,12</sup>, it may be necessary to use other methods to stimulate axon regeneration<sup>13</sup>. Future studies should investigate how best to facilitate integration of regenerated axons into local circuits and to harness the potential of anatomical plasticity to restore multiple functions after SCI. ■

**Katherine Zukor and Zhigang He** are at the Kirby Program in Neuroscience, Children's Hospital Boston, Boston, Massachusetts 02115, USA.  
e-mails: [katherine.zukor@childrens.harvard.edu](mailto:katherine.zukor@childrens.harvard.edu); [zhigang.he@childrens.harvard.edu](mailto:zhigang.he@childrens.harvard.edu)

1. Alilain, W. J., Horn, K. P., Hu, H., Dick, T. E. & Silver, J. *Nature* **475**, 196–200 (2011).
2. Feldman, J. L. & Del Negro, C. A. *Nature Rev. Neurosci.* **7**, 232–241 (2006).
3. Silver, J. & Miller, J. H. *Nature Rev. Neurosci.* **5**, 146–156 (2004).
4. Bradbury, E. J. *et al.* *Nature* **416**, 636–640 (2002).
5. David, S. & Aguayo, A. J. *Science* **214**, 931–933 (1981).
6. Houle, J. D. *et al.* *J. Neurosci.* **26**, 7405–7415 (2006).
7. Jones, B. W. & Marc, R. E. *Exp. Eye Res.* **81**, 123–137 (2005).
8. Goshgarian, H. G. *J. Appl. Physiol.* **94**, 795–810 (2003).
9. Pizzorusso, T. *et al.* *Science* **298**, 1248–1251 (2002).
10. García-Álías, G., Barkhuysen, S., Buckle, M. & Fawcett, J. W. *Nature Neurosci.* **12**, 1145–1151 (2009).
11. Richardson, P. M., McGuinness, U. M. & Aguayo, A. J. *Brain Res.* **237**, 147–162 (1982).
12. Benfey, M. & Aguayo, A. J. *Nature* **296**, 150–152 (1982).
13. Sun, F. & He, Z. *Curr. Opin. Neurobiol.* **20**, 510–518 (2010).



## 50 Years Ago

*Mathematical Methods in the Theory of Queueing.* By A. Y. Khintchine — This beautiful little book breathes reason and modesty from cover to cover ... A particularly pleasing feature is the way in which the results are developed; the mathematics is 'done', not merely indicated, and nowhere does the author state a result and refer the reader elsewhere for the details. Thus there is a satisfying aspect of completion about the exposition ... The theory of queues has undergone considerable development in recent years. Some mathematicians think the development has gone too far. Whether this is so or not the book under review will serve to show that the phenomenon of queueing represents another human experience which has bowed to the forces of applied mathematics; the concepts that have been built around this experience have proved to be of the right kind, and sufficient in number, for the mathematical development to go 'with a bang'.  
**From Nature 15 July 1961**

## 100 Years Ago

We published recently (June 29) a short article on the progress of radiography in medical diagnosis, and alluded in particular to the work of the staff at Guy's Hospital in their investigation of pathological conditions of the intestine. In this connection we note the appearance of a new paper ... by Dr. A. C. Jordan, medical radiographer to Guy's Hospital ... in which he shows that it is often possible to detect duodenal obstruction by the X-ray method after giving the patient a bismuth meal. Diseases of the duodenum are often extremely obscure, and this new method of diagnosing the condition will be welcomed both by the medical profession and the sufferers from such complaints.  
**From Nature 13 July 1911**

# Gentle measurement

**Ideally, measurement of the energy state of a single atom would set the atom to the measured state without affecting any of its other properties. This goal has now been achieved with the assistance of a small optical cavity. SEE LETTER P.210**

PETER MAUNZ

In classical physics, a measurement can in principle be carried out with unlimited precision without affecting the system being measured. In quantum physics, by contrast, every measurement that reveals information about a quantum system necessarily exerts a back-action on the system; this effect is also known as the collapse of the wavefunction. However, most measurements performed in the laboratory lead to a much larger back-action than is imposed by quantum theory.

On page 210 of this issue, Volz *et al.*<sup>1</sup> describe how an optical cavity can allow a single atom to be measured with essentially only the back-action required by quantum theory. Their result not only deepens our understanding of the boundary between quantum and classical physics, but is also a step towards making atom-based quantum-information processing a reality<sup>2</sup>.

Consider a single atom with two long-lived energy states. The atom can be either in one of these energy states or — and this is a property unique to quantum theory — in both states simultaneously. In an ideal ‘von Neumann’ measurement<sup>3,4</sup> of the energy state, the atom will be found in one of the two energy states.

Furthermore, the measurement back-action will set the atom to the measured state. Any subsequent measurement of the energy state will yield the same result. Besides setting the atom to the measured state, no further back-action is mandated by quantum theory. Although it should be possible to measure the energy state of the atom with no additional back-action involved in the process, and thus without energy exchange between the atom and the measurement probe, such an ideal measurement has not been achieved. Volz and colleagues’ study<sup>1</sup> offers the means to do just this.

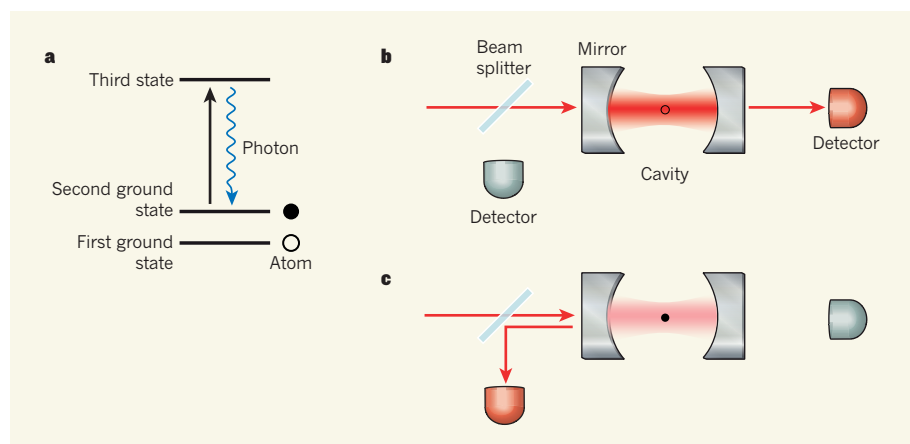
Previously, state-of-the-art measurements of the energy state of single trapped atoms were performed by fluorescence detection<sup>5,6</sup>. In this method, light is shone on the atom to excite it from one of its two ground states — say, the ‘bright’ state — into a third state, whereupon it spontaneously emits a photon and returns to the original state (Fig. 1a). With this closed optical-transition system, an atom in the bright state will absorb and emit (scatter) many photons. An atom in the other state, the ‘dark’ state, is not excited by the incident light and remains dark. Collecting and detecting the scattered photons allows the quantum state of the atom to be determined with high fidelity<sup>7</sup>. However, the large number of scattered photons can heat the

atom, and might leave it in a state other than its two ground states. Theoretical investigations<sup>8</sup> also show that detecting the state of an atom using a single-pass laser beam is always accompanied by spontaneous scattering of photons.

To overcome these limitations, Volz and colleagues<sup>1</sup> used a small optical cavity, an arrangement of two highly reflective mirrors that allows light to bounce back and forth between them (Fig. 1b,c). If a multiple of the half-wavelength of the laser light impinging on the cavity matches the distance between the mirrors, the cavity becomes nearly transparent to the light. By contrast, if this condition is not met, most light is reflected. The authors’ microscopic cavity, which was made of highly reflective coated optical fibres, can reflect a photon more than 40,000 times on average before it is lost. This microscopic cavity with its high-reflectivity mirrors enabled the authors to reach a regime in which even a single atom can strongly shift the wavelength at which light resonates in the cavity.

In this set-up, an atom in the bright state, with an optical transition that is resonant with the cavity, shifts the cavity’s resonance frequency, and most of the incoming light is reflected (Fig. 1c). In the dark state, the atom does not ‘see’ the light and consequently does not change the resonance frequency of the cavity, which remains transparent to the incident light (Fig. 1b). What’s more, in both states the atom scatters hardly any photons. By separately detecting photons reflected from and transmitted through the cavity, the authors could measure the quantum state of the atom<sup>9</sup> without energy exchange being incurred. Volz and colleagues<sup>1</sup> determined the state of a single atom with more than 90% fidelity and with spontaneous scattering of less than 0.2 photons on average — therefore, the atom is subjected to almost no heating.

The knowledge about the atom’s quantum state that the experiment can reveal is limited by the quantum efficiency of the instruments used to detect the light. Nevertheless, by means of a feature known as the quantum Zeno effect<sup>10</sup>, the authors were able to quantify the back-action exerted on the atom and thereby determine the average time between measurements. The atom is first prepared in one of the two ground states, for example in the bright state. Before the state is measured, an incident microwave pulse induces transitions between the two ground states. After the pulse, the atom is always found in the other state — the dark state, in this example. If the state of the atom is measured after a much shorter time than the duration of the pulse, there is a high probability of finding the atom in the original state: in this case, the measurement back-action resets the atom to the original state. Decreasing the time between several state measurements strongly increases the probability of finding the atom in the original state. As a consequence, the atom can be frozen in the initial quantum state by measuring it very



**Figure 1 | Measuring the quantum state of a single atom<sup>1</sup>.** **a**, An atom that can exist in two ground states can be in either of these states or in both states simultaneously. Incident laser light of appropriate frequency (not shown) can excite the atom from the second ground state to a third, higher-energy state, whereupon it returns to the original state while emitting a photon. **b**, In an optical cavity formed by two highly reflective mirrors, an atom in the first ground state is not influenced by the incoming light, nor does it change the properties of the cavity. Most light is transmitted through the cavity to the detector located after the second mirror. **c**, In the second ground state, the atom shifts the cavity’s resonance frequency, and most of the light is reflected at the first mirror. A partly transparent mirror (beam splitter) is used to direct the light reflected from the cavity to a photodetector.



often — this is the quantum Zeno effect.

Volz and colleagues employ this effect to determine the frequency with which the atomic state is reset, and thus the frequency with which measurements of the atomic state are performed. They show that an average of two photons incident on the cavity is sufficient for a measurement of the internal state of the atom, setting the atomic state to either the bright or dark state. This result quantifies the measurement back-action and demonstrates the potential of cavity-assisted state detection of single atoms.

Volz and co-workers' experiment<sup>1</sup> furthers our understanding of the quantum-measurement process and the interactions between a quantum system and its environment. It can be used to execute fast quantum-state measurements of atoms that deliver a result for each atom. The ability to make such measurements

is essential for fundamental tests of quantum mechanics<sup>11</sup>. Measurement of the quantum state without heating while simultaneously allowing preparation of the atom in one of the states is also an important tool for creating a neutral-atom-based quantum computer<sup>2</sup>.

The fidelity of quantum-state detection and the spontaneous photon-scattering rate achieved in the current experiment are still limited by technical imperfections in the cavity<sup>1</sup>. An improved cavity, or one embedded in an interferometer<sup>12</sup>, would open the path to quantum-state detection of trapped molecules without the need to use closed optical transitions<sup>13</sup>. ■

**Peter Maunz** is in the Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina 27708-0291, USA.  
e-mail: peter.maunz@duke.edu

1. Volz, J., Gehr, R., Dubois, G., Estève, J. & Reichel, J. *Nature* **475**, 210–213 (2011).
2. Ladd, T. D. *et al.* *Nature* **464**, 45–53 (2010).
3. von Neumann, J. *Mathematische Grundlagen der Quantenmechanik* (Springer, 1932).
4. Schlosshauer, M. *Rev. Mod. Phys.* **76**, 1267–1305 (2005).
5. Sauter, T., Neuhauser, W., Blatt, R. & Toschek, P. E. *Phys. Rev. Lett.* **57**, 1696–1698 (1986).
6. Bergquist, J. C., Hulet, R. G., Itano, W. M. & Wineland, D. J. *Phys. Rev. Lett.* **57**, 1699–1702 (1986).
7. Myerson, A. H. *et al.* *Phys. Rev. Lett.* **100**, 200502 (2008).
8. Hope, J. J. & Close, J. D. *Phys. Rev. A* **71**, 043822 (2005).
9. Bochmann, J. *et al.* *Phys. Rev. Lett.* **104**, 203601 (2010).
10. Itano, W. M., Heinzen, D. J., Bollinger, J. J. & Wineland, D. J. *Phys. Rev. A* **41**, 2295–2300 (1990).
11. Simon, C. & Irvine, W. T. M. *Phys. Rev. Lett.* **91**, 110405 (2003).
12. Long, R. *et al.* *Phil. Trans. R. Soc. Lond. A* **361**, 1375–1389 (2003).
13. Jones, K. M., Tiesinga, E., Lett, P. D. & Julienne, P. S. *Rev. Mod. Phys.* **78**, 483–535 (2006).

## CARDIOVASCULAR DISEASE

# Several small shocks beat one big one

**Life-threatening abnormalities in the electrical rhythm of the heart are usually treated with the application of a large electric shock. An approach involving a significantly smaller shock energy may be equally effective. SEE LETTER P.235**

RICHARD A. GRAY & JOHN P. WIKSWO

Electricity can kill, but it can also revive individuals experiencing life-threatening cardiac arrhythmias when it is applied through external or implanted cardiac defibrillator devices. Sudden cardiac death is the result of fibrillation — irregular contractions of the heart muscle caused by multiple electrical waves rotating throughout the heart<sup>1</sup>. Although these 'rotor' waves are unstable, their spatial and temporal organization provides hope that defibrillation-shock energies can be significantly decreased through the design of better devices. Then, not only may the size and cost of defibrillator devices be reduced, but so may the pain experienced by patients during a shock<sup>2</sup>. On page 235 of this issue, Luther *et al.*<sup>3</sup> describe a significant advance in this direction by showing the efficacy of a sequence of low-energy electric shocks.

Fibrillation is not driven by a particular heart region, but is sustained by rotor waves across the heart. This means that halting fibrillation requires an approach that affects the entire heart. Traditionally, this is done using two widely spaced metal electrodes to deliver a high-energy shock, which creates a large electric field throughout the heart. The current flowing between the two electrodes

follows the paths of least resistance, and so crosses cell membranes, thereby changing — both increasing and decreasing — the electrical potential across the membranes ( $V_m$ ) in affected regions and creating what are known as virtual electrodes<sup>4</sup>. A successful defibrillation shock quickly restores  $V_m$  to resting levels across the heart (Fig. 1a,b).

One approach to reducing the strength of the defibrillation shock is to use several metal electrodes placed throughout the heart<sup>5,6</sup>. At each of the electrode locations, a propagating wave induced by electric stimulation can alter the dynamics of rotor waves, if the pacing rate is faster than the fibrillation frequency (overdrive pacing)<sup>7,8</sup>. This approach, however, has not proven very effective in the clinic<sup>9,10</sup>, partly because of the challenges involved in implanting multiple electrodes in the heart.

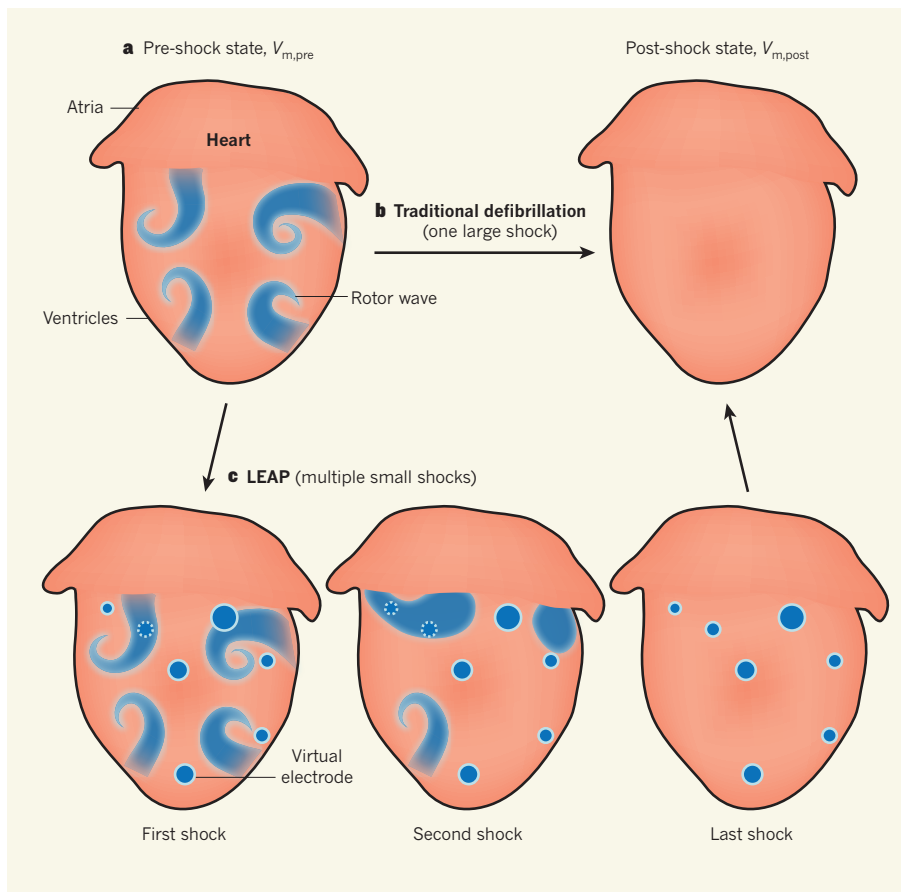
Virtual electrodes ( $\Delta V_m$ ) can be generated in the heart by the interaction between the electric field and the tissues of the heart and thorax. They are affected by membrane resistance and capacitance; by tissue conductivity; by the orientation of muscle fibres; and by the detailed geometry of electrical inhomogeneities such as those seen at the boundaries between the heart and lungs, or between the heart tissue, blood vessels and scar tissue. The equations representing these phenomena are

well understood and involve one of the four fundamental forces of physics — electromagnetism. Nonetheless, rigorous simulations of the effect of an electric field on the heart are possible only if all the relevant geometry, conductivities and boundary conditions can be accurately described.

To make matters even more complicated, successful defibrillation is not a result of just the interaction between the applied electric field and the tissue. It is also influenced by the pre-shock state of the transmembrane potential ( $V_{m,pre}$ ) and the stability of the rotor waves. The combined effects of tissue heterogeneity and the unpredictable nature of fibrillation mean that the outcome of the shock treatment varies between individuals and even in the same individual during a cardiac arrest. For traditional defibrillation, the post-shock transmembrane potential ( $V_{m,post}$ ) resulting from a single, brief, high-energy shock determines whether the shock is successful, and this depends largely on whether any sustainable rotor waves remain after the shock (Fig. 1b).

Although any spatial heterogeneity in the conductivity of the normal or diseased heart is known to contribute to the generation of virtual electrodes<sup>11–13</sup>, it is not known which heterogeneities are most relevant for re-establishing synchronous contractions. Luther *et al.*<sup>3</sup> show that the coronary arteries can have a large role in determining the precise spatial patterns of virtual electrodes throughout the heart — a finding that is also supported by another study<sup>14</sup>.

Armed with the knowledge that, with increasing shock strength, the tissue regions experiencing virtual electrodes increase and the time taken to excite the entire heart (the global activation time) decreases<sup>15</sup>, Luther and colleagues build on an elegant scaling argument<sup>16</sup>. Their results relate the global activation time obtained in their experiments to the size distribution of the branching structure



**Figure 1 | Subtle defibrillation.** **a**, During cardiac fibrillation, the pre-shock electrical state of the heart (the transmembrane potential throughout the heart;  $V_{m,pre}$ ) features numerous unstable, electrical rotating waves (rotor waves). **b**, Traditional defibrillation works to rapidly bring the post-shock state ( $V_{m,post}$ ) to resting levels by means of a single, large electrical shock. **c**, Luther *et al.*<sup>3</sup> describe low-energy antibrillation pacing (LEAP), in which multiple low-strength shocks are applied to generate many virtual electrodes across the heart. These virtual electrodes excite the heart tissue, and the waves generated by each pulse of electricity propagate away from the electrodes, interacting with the rotor waves to eventually terminate fibrillation.

of the coronary vasculature. The branching structure was computed from polymer casts obtained from the same experiments. The authors propose that during a shock, the specific geometry (size, orientation and so on) of the blood-filled coronary vasculature, and its difference in electrical conductivity from that of the surrounding heart, lead to the formation of virtual electrodes throughout much of the heart. Indeed, Luther *et al.* show that shock strengths significantly lower than those required for traditional defibrillation can produce many virtual electrodes throughout the heart of beagle dogs, and that these electrodes generate numerous propagating waves without the need for multiple metal electrodes.

Through their *in vitro* and *in vivo* animal experiments, the researchers<sup>3</sup> show that the application of multiple, brief shocks significantly reduces the energy required for defibrillation by launching numerous propagating waves from many virtual electrodes across the heart (Fig. 1c). Surprisingly, the interval between the application of each low-energy antibrillation pacing (LEAP) stimulus was longer than the average rotor period (underdrive pacing).

These intriguing findings lead to equally intriguing questions, an obvious one being whether this phenomenon will translate to humans. Moreover, the curious researcher might wonder whether even lower defibrillation energies — which are always desirable — could be used. However, there may be a theoretical minimum energy for various reasons, including the fact that a minimum electric field (around 1 volt per centimetre) is required to excite cardiac tissue<sup>17</sup>. In a given patient, the success of LEAP defibrillation will depend on the relationship of rotor-wave density and location and the spatial distribution and size of heterogeneities in tissue conductivity.

In light of Luther and co-workers' study, it seems prudent to factor coronary vasculature into whole-heart simulations of defibrillation. But deciding how to do this appropriately requires a detailed understanding of the geometry, conductivity and path of the electrical current near blood vessels. Such details can vary widely between individuals.

Although provocative, the new work<sup>3,14</sup> does not directly show the exact mechanism

involved, nor does it outline the precise experimental pattern of virtual-electrode polarization resulting from vasculature-induced heterogeneities. Factors to explore include the effects of vessel shape and of vessel-wall and blood conductivities on the generation of virtual electrodes<sup>18</sup>. That said, it is exciting that LEAP can reduce the defibrillation threshold for both atrial and ventricular fibrillation *in vitro* and can terminate atrial fibrillation *in vivo* (by means of coil electrodes inside the heart). Indeed, LEAP is an important development showing that a significant decrease in the required shock strength results from a combination of dynamic control and the interaction of the electric field with the heart structure — purportedly, the coronary vasculature. ■

**Richard A. Gray** is in the Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, Maryland 20993-0002, USA. **John P. Wikswo** is in the Departments of Biomedical Engineering, of Molecular Physiology & Biophysics, and of Physics & Astronomy, Vanderbilt University, Nashville, Tennessee 37235, USA. e-mails: richard.gray@fda.hhs.gov; john.wikswo@vanderbilt.edu

- Gray, R. A., Pertsov, A. M. & Jalife, J. *Nature* **392**, 75–78 (1998).
- Dosdall, D. J. & Ideker, R. E. *Heart Rhythm* **4**, S51–S56 (2007).
- Luther, S. *et al.* *Nature* **475**, 235–239 (2011).
- Wikswo, J. P., Lin, S. F. & Abbas, R. A. *Biophys. J.* **69**, 2195–2210 (1995).
- Pak, H.-N. *et al.* *Am. J. Physiol. Heart Circ. Physiol.* **285**, H2704–H2711 (2003).
- Hosfeld, V. D. *et al.* *J. Biol. Phys.* **33**, 145–153 (2007).
- Allessie, M. *et al.* *Circulation* **84**, 1689–1697 (1991).
- Davidenko, J. M. *et al.* *Circ. Res.* **77**, 1166–1179 (1995).
- Ditto, W. L. *et al.* *Int. J. Bifurc. Chaos* **10**, 593–601 (2000).
- Kok, L. C. & Ellenbogen, K. A. *Cardiol. Clin.* **22**, 71–86 (2004).
- Fishler, M. G. *J. Cardiovasc. Electrophysiol.* **9**, 384–394 (1998).
- Sobie, E. A., Susil, R. C. & Tung, L. *Biophys. J.* **73**, 1410–1423 (1997).
- Trayanova, N. A., Roth, B. J. & Malden, L. J. *IEEE Trans. Biomed. Eng.* **40**, 899–908 (1993).
- Bishop, M. J. *et al.* *IEEE Trans. Biomed. Eng.* **57**, 2335–2345 (2010).
- Maleckar, M. M. *et al.* *Am. J. Physiol. Heart Circ. Physiol.* **295**, H1626–H1633 (2008).
- Pumir, A. *et al.* *Phys. Rev. Lett.* **99**, 208101 (2007).
- Ideker, R. E., Zhou, X. & Knisley, S. B. *Pacing Clin. Electrophysiol.* **18**, 512–525 (1995).
- Cartee, L. A. & Plonsey, R. *IEEE Trans. Biomed. Eng.* **39**, 76–85 (1992).

#### CORRECTION

An erroneous Figure 1 appeared in the News & Views article “Materials science: Graphene moiré mystery solved?” by Allan H. MacDonald & Rafi Bistritzer (*Nature* **474**, 453–454; 2011), in that the lattices depicted were not honeycombs. The correct figure is now in place in the online version of the article.



# Collective synthesis of natural products by means of organocascade catalysis

Spencer B. Jones<sup>1</sup>, Bryon Simmons<sup>1</sup>, Anthony Mastracchio<sup>1</sup> & David W. C. MacMillan<sup>1</sup>

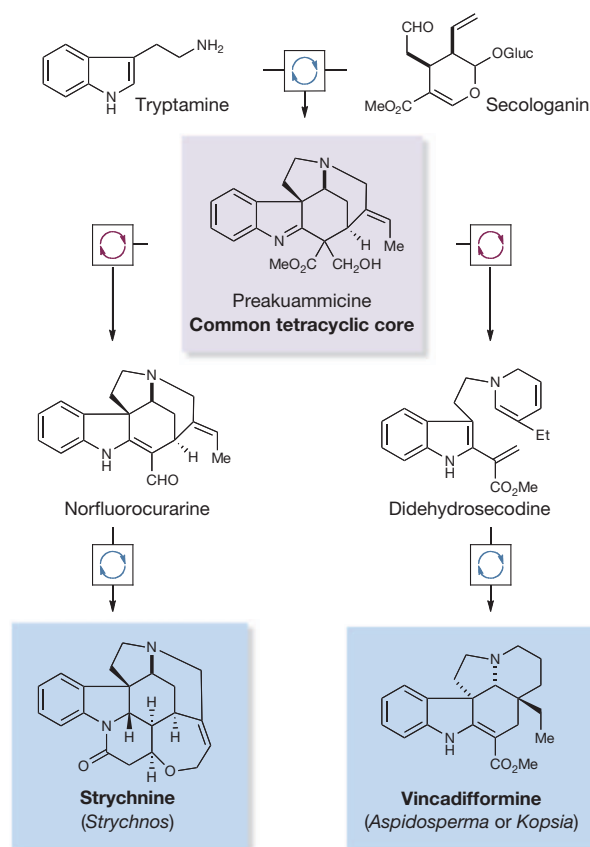
Organic chemists are now able to synthesize small quantities of almost any known natural product, given sufficient time, resources and effort. However, translation of the academic successes in total synthesis to the large-scale construction of complex natural products and the development of large collections of biologically relevant molecules present significant challenges to synthetic chemists. Here we show that the application of two nature-inspired techniques, namely organocascade catalysis and collective natural product synthesis, can facilitate the preparation of useful quantities of a range of structurally diverse natural products from a common molecular scaffold. The power of this concept has been demonstrated through the expedient, asymmetric total syntheses of six well-known alkaloid natural products: strychnine, aspidospermidine, vincadifformine, akuammicine, kopsanone and kopsinine.

The field of natural product total synthesis has evolved dramatically over the past 60 years owing substantially to the efforts of strategy-based organic chemists along with remarkable improvements in our bond-forming capabilities. However, two critical challenges that remain for this field are those of translating laboratory-level academic success in total synthesis to the large-scale assembly of biologically important molecules<sup>1</sup> and building large collections of natural product families (or their analogues) for use as biological probes or in medicinal chemistry<sup>2</sup>.

To address these two fundamental challenges in small-molecule synthesis, we have been inspired by the strategies used in nature to solve these problems. For example, as chemists, we typically use 'stop-and-go' synthetic protocols, whereby individual transformations are conducted as stepwise processes punctuated by the isolation and purification of intermediates at each stage of the sequence<sup>3</sup>. By contrast, in nature the rapid conversion of simple starting materials to complex molecular scaffolds is accomplished through the use of transformation-specific enzymes, which mediate a continuous series of highly regulated catalytic cascades in what amounts to be a highly efficient 'biochemical assembly line'<sup>4,5</sup>. Moreover, biosynthesis in nature provides an appealing alternative to the traditional 'single-target' approach to chemical synthesis in that it typically involves the construction of natural product collections through the assembly of a common intermediate (Fig. 1).

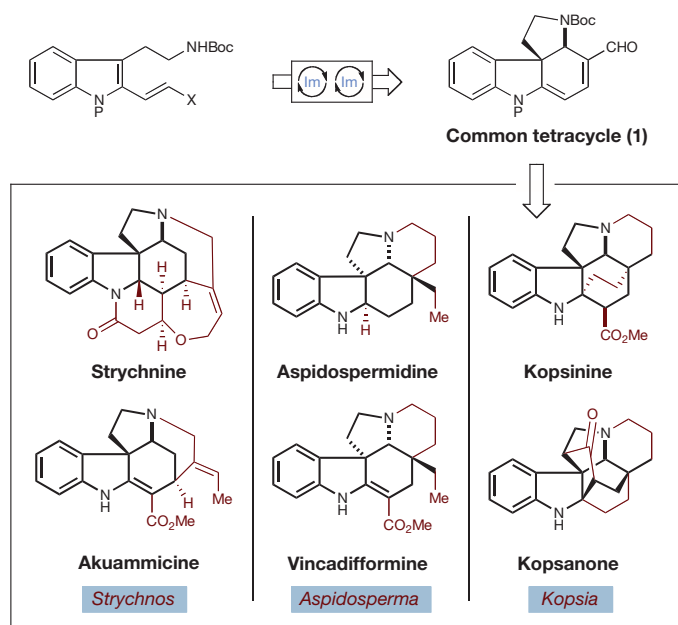
## Design plan

We recently sought to develop a novel asymmetric approach to total synthesis based on the application of these two nature-inspired concepts, namely collective total synthesis and organocascade catalysis<sup>3,4</sup>. Whereas syntheses targeting an advanced core structure applicable to the synthesis of closely related natural products within a family have been frequently reported<sup>6</sup>, much less common is the preparation of an intermediate endowed with functionality amenable to the preparation of structurally diverse natural products in different families—a strategy we term collective total synthesis<sup>7,8</sup>. We predicted that the hybridization of the strategies of collective natural product synthesis and enantioselective organocascade catalysis<sup>9</sup> should rapidly give access to useful quantities of single-enantiomer natural product collections or families with unprecedented levels of ease and efficiency.



**Figure 1 | Cascade catalysis in biosynthesis.** In nature, transform-specific enzymes in continuous catalytic cascades rapidly produce common biosynthetic intermediates and natural products. Preakuammicine serves as a biosynthetic precursor to a range of structurally diverse members of the *Strychnos*, *Aspidosperma* and *Kopsia* alkaloid families, including strychnine and vincadifformine. Et, ethyl; Gluc, glucose; Me, methyl.

<sup>1</sup>Merck Center for Catalysis at Princeton University, Princeton, New Jersey 08544, USA.



**Figure 2 | Collective natural product synthesis: nature-inspired application of cascade catalysis.** Synthesis of six structurally diverse *Strychnos*, *Aspidosperma* and *Kopsia* alkaloids is expected to proceed from a common intermediate tetracycle prepared by means of organocascade catalysis. Boc, tert-butoxycarbonyl; Im, iminium catalysis.

To demonstrate this approach, we targeted for total synthesis six well-known, structurally complex members of the *Strychnos*, *Aspidosperma* and *Kopsia* families of alkaloids, in particular strychnine, the best-known member of the *Strychnos* family, which has been the focus of intense synthetic interest for over 50 years. Moreover, this target has served as a key metric for the success of our nature-inspired dual techniques<sup>10,11</sup>.

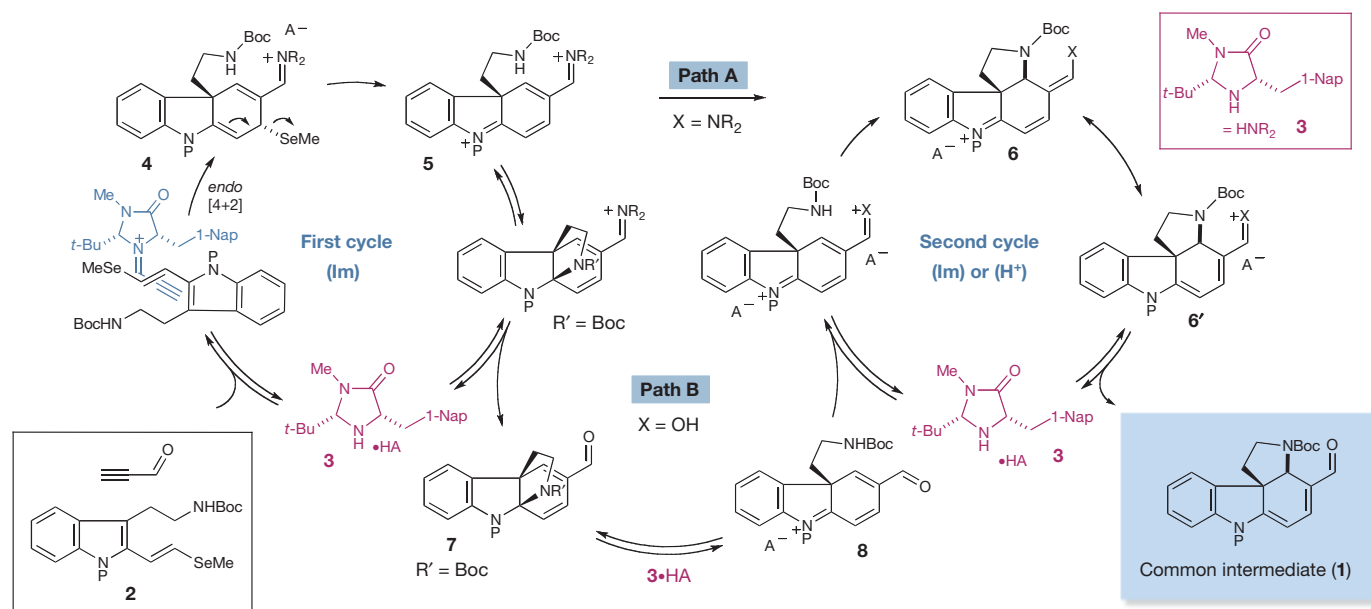
Importantly, strychnine is believed to share a biosynthetic precursor with a range of prominent *Strychnos*, *Aspidosperma* and *Kopsia* alkaloids<sup>12</sup>. This common intermediate, preakuammicine, arises biosynthetically through a controlled enzymatic cascade involving a coupling of the precursors tryptamine and secologanin,

followed by skeletal rearrangement (Fig. 1). Thus, in a prime example of natural economy, a single intermediate is exploited in the construction of a diverse collection of complex molecular products.

We expected the preparation of intermediate **1**, which incorporates the requisite functionality for expedient conversion to each of the target natural products (Fig. 2), to be a central element of our design strategy. The key tetracyclic precursor **1** would itself be accessed through a one-flask, asymmetric Diels–Alder/elimination/conjugate addition organocascade sequence commencing with a simple tryptamine-derived substrate<sup>13</sup>.

In detail, we hoped that exposure of the 2-(vinyl-1-selenomethyl) tryptamine system **2** to propynal in the presence of an imidazolidinone catalyst (**3**) would set in motion a Diels–Alder [4 + 2] addition<sup>13</sup> (Fig. 3). As a key element of enantiocontrol, the acetylenic functionality of the catalyst-bound propynal would be expected to partition away from the bulky tert-butyl (*t*-Bu) group, leaving the naphthyl group effectively to shield the bottom face of the reacting alkyne. The activated dienophile would then undergo *endo*-selective Diels–Alder cycloaddition with the substituted 2-vinyl indole **2**. By contrast with our previous studies using methyl-sulphide-substituted 2-vinyl indoles, as demonstrated in our synthesis of minfiensine<sup>13</sup>, we felt a conceptually unique series of cascade cycles might arise through the incorporation of organoselenide substitution on the diene, leading to a novel, architecturally complex system. More specifically, following the Diels–Alder cycloaddition, the cycloadduct **4** would be poised to undergo facile  $\beta$ -elimination of methyl selenide to furnish the unsaturated iminium ion **5**. We proposed this change in mechanism owing to the higher propensity of selenides to undergo  $\beta$ -elimination in comparison with sulphides<sup>14</sup>.

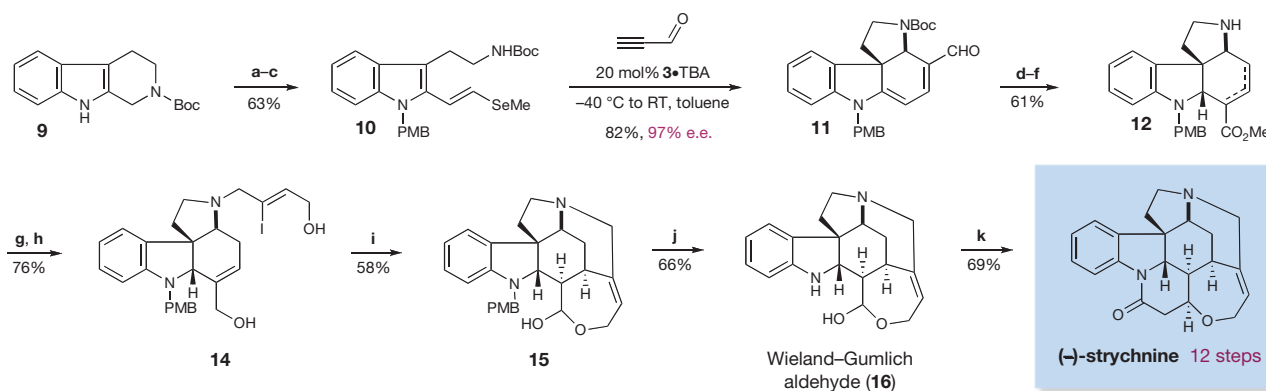
In the second cycle, iminium-catalysed 5-*exo*-heterocyclization of the pendant carbamate was expected to occur at the  $\delta$ -position to the indolinium ion (**5**  $\rightarrow$  **6**; Fig. 3, path A, X = NR<sub>2</sub>) to deliver, following hydrolysis, the enantioenriched spiroindoline core (**1**). We considered the possibility of an alternative second cycle wherein iminium **5** might undergo facile cyclization at the indoline carbon to generate pyrroloindoline **7** transiently (Fig. 3, path B). In this situation, we recognized that amine or Brønsted acid catalysis might thereafter induce the necessary 5-*exo*-heterocyclization of the pendant carbamate to also furnish **6**. Importantly, either cascade sequence could allow for the rapid and enantioselective production of the complex tetracyclic spiroindoline



**Figure 3 | Proposed mechanism of organocascade cycles for the generation of a common tetracyclic intermediate (1).** An organocascade reaction between 2-vinyl indole **2** and propynal is expected to proceed through an

organocatalytic Diels–Alder/ $\beta$ -elimination/amine conjugate addition sequence along path A, involving iminium ion catalysis, or path B, involving Brønsted acid catalysis. 1-Nap, 1-naphthyl; SeMe, selenomethyl.





**Figure 4 | Twelve-step enantioselective total synthesis of (–)-strychnine.** Reagents and conditions are as follows. **a**, NaH, PMBCl, dimethylformamide (DMF), 0 °C. PMB, *para*-methoxybenzyl. **b**, SeO<sub>2</sub>, dioxane, H<sub>2</sub>O, 100 °C. **c**, (EtO)<sub>2</sub>P(O)CH<sub>2</sub>SeMe, 18-crown-6, potassium bis(trimethylsilyl)amide (KHMDs), tetrahydrofuran (THF), –78 °C to room temperature (RT, 23 °C). e.e., enantiomeric excess. **d**, (Ph<sub>3</sub>P)<sub>3</sub>RhCl, toluene, PhCN, 120 °C. **e**, COCl<sub>2</sub>,

Et<sub>3</sub>N, toluene, –45 °C to RT, then MeOH, –30 °C to RT. **f**, DIBAL-H, CH<sub>2</sub>Cl<sub>2</sub>, –78 °C to RT, then trifluoroacetic acid (TFA), –78 °C to RT. **g**, 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU), K<sub>2</sub>CO<sub>3</sub>, DMF, (Z)-4-bromo-3-iodobut-2-enyl acetate (**13**), RT. **h**, DIBAL-H, CH<sub>2</sub>Cl<sub>2</sub>, –78 °C. **i**, 25 mol% Pd(OAc)<sub>2</sub>, Bu<sub>4</sub>NCl, NaHCO<sub>3</sub>, EtOAc, RT. **j**, PhSH, TFA, 45 °C. **k**, NaOAc, Ac<sub>2</sub>O, AcOH, malonic acid, 120 °C. Ac, acetyl.

1 from simple tryptamine-derived and ynal substrates in a single operation.

## Experimental results

The feasibility of the proposed organocascade sequence was first evaluated in the context of a total synthesis of strychnine. The requisite 2-vinyl indole **10** was prepared in three steps from **9** according to the standard procedures outlined in Fig. 4 (ref. 15). The crucial organocascade addition–cyclization was accomplished with the use of 1-naphthyl-substituted imidazolidinone catalyst **3** in the presence of 20 mol% tribromoacetic acid (TBA) co-catalyst, forming the complex spiroindoline **11** in 82% yield and with excellent levels of enantioinduction (97% e.e.). Notably, we have now gathered evidence that path B of the cascade sequence is operational. Specifically, when the same protocol was performed in the presence of stoichiometric catalyst at –78 °C and quenched after 10 min with Et<sub>3</sub>N, an 84% yield of pyrroloindoline **7** (protecting group, PMB) was obtained. Moreover, exposure of pyrroloindoline **7** (protecting group, PMB) to catalytic **3**·TBA and, separately, *N*-methyl **3**·TBA (incapable of undergoing iminium formation) facilitates the conversion to the spiroindoline **11** at comparable rates.

The product of the key organocascade sequence, tetracyclic spiroindoline **11**, was advanced to strychnine in only eight additional steps (Fig. 4). In detail, decarbonylation was achieved through the use of Wilkinson's catalyst. Subsequent treatment with phosgene and methanol<sup>16</sup> served to introduce a carbomethoxy group at the dieneamine α-position. Next, on exposure to DIBAL-H, the enamine unsaturation was reduced and the requisite tertiary indoline stereocentre was installed to provide the unsaturated ester **12** (existing as an inconsequential mixture of alkene isomers) in 62% overall yield for the three steps. We converted intermediate **12** to vinyl iodide **14** through a two-step 76%-yield protocol involving allylation with the substituted allyl bromide **13** and concomitant alkene isomerization, followed by DIBAL-H-mediated reduction of both ester functionalities.

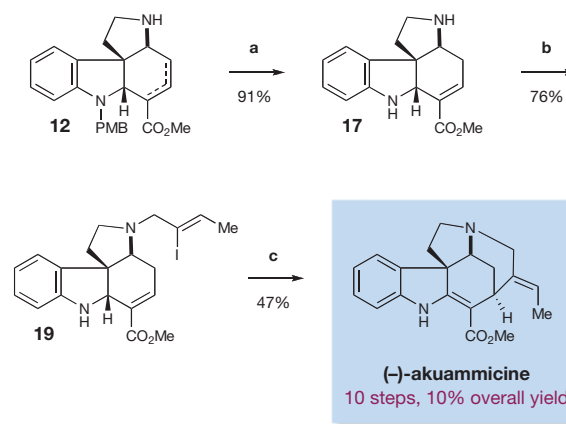
As a second key step, we predicted direct conversion of the vinyl iodide **14** to the protected Wieland–Gumlich aldehyde **15** through a cascade Jeffery–Heck cyclization/lactol formation sequence<sup>17</sup>. Insertion of palladium into the vinyl iodide and subsequent carbopalladation would forge the six-membered ring and produce an alkyl palladium intermediate, which would then undergo β-hydride elimination to provide an enol that would rapidly engage in lactol formation with the proximal alcohol-bearing side chain.

Following extensive investigations, we discovered an optimized set of conditions to allow successful Jeffery–Heck cyclization/lactol

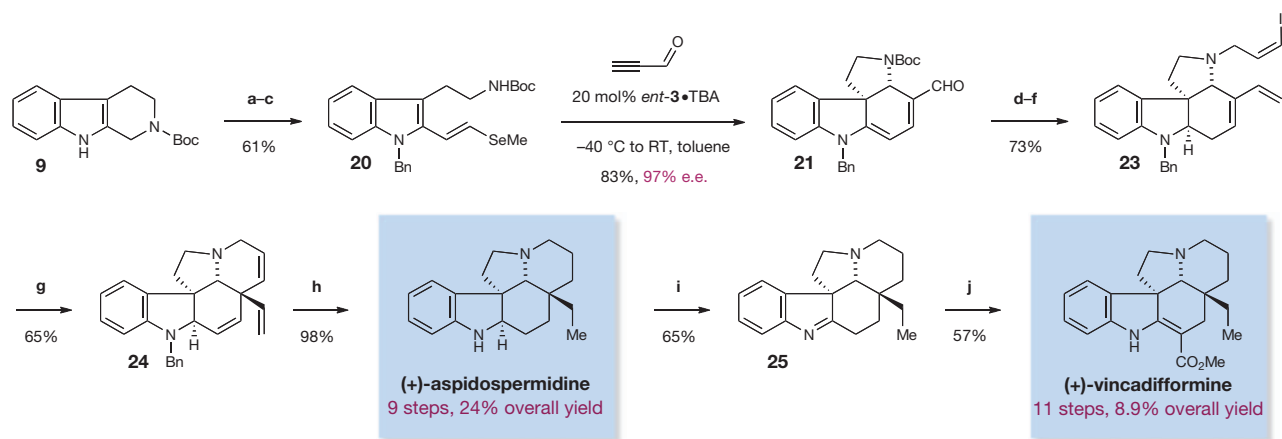
formation, using vinyl iodide **14** to form the PMB-protected Wieland–Gumlich aldehyde **15** in 58% yield<sup>18</sup>. Early studies showed that the PMB protecting group was critical in facilitating regioselective β-hydride elimination away from the indoline ring methine (and productively towards the alcohol). Such a high degree of regiocontrol presumably arises from the allylic strain that accompanies formation of the *N*-PMB-substituted enamine, a destabilizing element that is absent from the corresponding enol formation step.

Finally, the synthetic sequence was completed by TFA-mediated removal of the PMB group<sup>13</sup> to furnish the Wieland–Gumlich aldehyde (**16**), which, on heating in a mixture of malonic acid, acetic anhydride and sodium acetate<sup>18</sup>, delivered enantioenriched (–)-strychnine in 12 steps and 6.4% overall yield from commercial materials. To our knowledge, this asymmetric organocascade-based sequence is the shortest route to enantioenriched strychnine that has been accomplished so far<sup>19–22</sup>.

We next turned to the construction of related alkaloids of the *Strychnos*, *Aspidosperma*, and *Kopsia* families based on the strategy of collective total synthesis outlined above. The total synthesis of (–)-akuammicine<sup>11</sup> (Fig. 5) was achieved starting with unsaturated ester **12** (prepared in the course of the synthesis of strychnine; see Fig. 4). Treatment of the PMB-protected spiroindoline **12** with TFA and thiophenol at 60 °C resulted in the cleavage of the PMB protecting group as well as isomerization of the alkene into conjugation with the



**Figure 5 | Ten-step enantioselective synthesis of (–)-akuammicine.** Reagents and conditions are as follows. **a**, TFA, PhSH, 60 °C. **b**, (Z)-1-bromo-2-iodobut-2-ene (**18**), K<sub>2</sub>CO<sub>3</sub>, DMF, RT. **c**, 20 mol% Pd(OAc)<sub>2</sub>, NaHCO<sub>3</sub>, Bu<sub>4</sub>NCl, MeCN, 65 °C.



**Figure 6** | Enantioselective total syntheses of (+)-aspidospermidine and (+)-vincadifformine. Reagents and conditions are as follows. **a**, NaH, DMF, BnBr, RT. **b**, SeO<sub>2</sub>, dioxane, H<sub>2</sub>O, 100 °C. **c**, (EtO)<sub>2</sub>P(O)CH<sub>2</sub>SeMe, 18-crown-6, KHMDS, THF, -78 °C to RT. *ent*-, *enantio*-. **d**, Ph<sub>3</sub>PCH<sub>3</sub>I, *n*-butyllithium,

THF, 0 °C, then AcOH, NaCNBH<sub>3</sub>, 0 °C. Bn, benzyl. **e**, TFA, CH<sub>2</sub>Cl<sub>2</sub>, RT. **f**, (Z)-3-bromo-1-iodoprop-1-ene (**22**), K<sub>2</sub>CO<sub>3</sub>, DMF, RT. **g**, (Ph<sub>3</sub>P)<sub>4</sub>Pd, Et<sub>3</sub>N, toluene, 80 °C. **h**, Pd(OH)<sub>2</sub>, H<sub>2</sub> (200 p.s.i.), MeOH, EtOAc, RT. **i**, CH<sub>2</sub>Cl<sub>2</sub>, DMSO, (COCl)<sub>2</sub>. **j**, *n*-butyllithium, NCCO<sub>2</sub>Me, THF, -78 °C to RT.

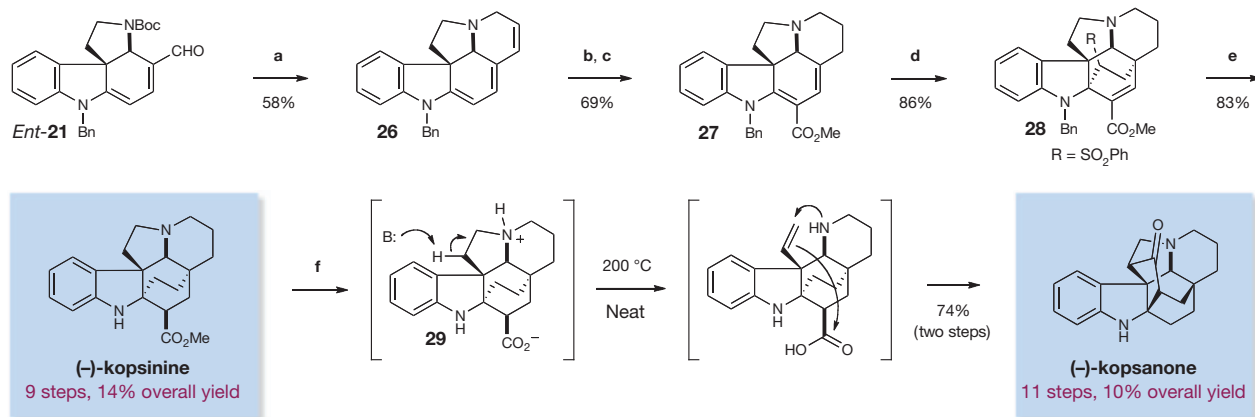
ester to give diamine **17** in 91% yield. Alkylation of the pyrrolidine nitrogen with functionalized allyl bromide **18** (ref. 23) gave vinyl iodide **19**, a precursor to akuammicine, by means of a Heck cyclization<sup>11</sup>. We expected insertion of palladium into the vinyl iodide followed by carbopalladation of the  $\alpha,\beta$ -unsaturated ester to give an alkyl palladium intermediate.  $\beta$ -hydride elimination would then furnish the natural product. In the event, treatment of vinyl iodide **19** with palladium acetate under Jeffery conditions<sup>24</sup> gave (-)-akuammicine, prepared in a total of ten steps and 10% overall yield.

The alkaloids aspidospermidine<sup>25</sup> and vincadifformine<sup>26</sup> are among the most highly sought *Aspidosperma* alkaloid targets. Aspidospermidine, in particular, has been the subject of extensive investigation, having been synthesized by over 30 research groups<sup>27,28</sup>. Application of our cascade catalysis approach allowed access to both of these alkaloids according to the following sequence. Conversion of the cascade product **21** into the vinyl iodide **23** was achieved in a three-step sequence as outlined in Fig. 6. After optimization, we found that a Heck cyclization of the vinyl iodide onto the tri-substituted double bond could be used to form triene **24** in 65% yield, thus completing the pentacyclic core of aspidospermidine. We achieved a simultaneous global hydrogenation/debenzylation of triene **24** using palladium hydroxide on carbon under hydrogen pressure, to afford (+)-aspidospermidine in nine linear steps and 23% overall yield (Fig. 6), which constitutes the shortest enantioselective synthesis of aspidospermidine reported so far<sup>29,30</sup>. Finally, we recognized the potential to synthesize (+)-vincadifformine from (+)-aspidospermidine by means of an oxidation and carbomethoxylation

sequence<sup>30</sup>. Indeed, Swern oxidation of aspidospermidine leads to the formation of imine **25**, which can be treated with *n*-butyllithium followed by methyl cyanofornate to obtain (+)-vincadifformine (Fig. 6) in 11 steps and 8.9% overall yield<sup>31</sup>.

We further extended our strategy of collective total synthesis to the synthesis of kopsinine<sup>32–35</sup> and the related compound kopsanone<sup>33,36</sup>. A unified approach to producing both alkaloids was implemented, allowing for a two-step conversion of kopsinine to kopsanone by means of a biomimetic thermocyclization<sup>33,37</sup> (Fig. 7). Kopsinine was synthesized by first treating *enantio*-**21** with trimethylsilyl iodide and then by vinyl triphenylphosphonium bromide to induce a deprotection/conjugate addition. Further treatment with KO<sup>t</sup>-Bu promoted a Wittig olefination to form the cyclic alkene found in triene **26**. An enamine  $\alpha$ -carbomethoxylation was then accomplished with phosgene-methanol<sup>16</sup> to afford an intermediate ester, which was selectively reduced to diene **27** by treatment with palladium on carbon (Pd/C) and H<sub>2</sub> in 69% over two steps.

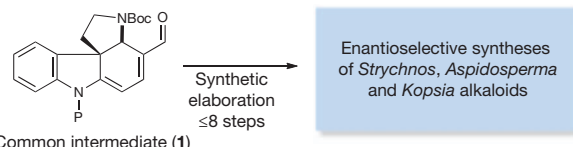
Dienes such as **27** can undergo [4 + 2] cycloadditions with a range of dienophiles, including vinyl sulfones<sup>33,38</sup>. We were able to obtain cycloadduct **28** in 83% yield through treatment of diene **27** with phenylvinyl sulfone in refluxing benzene. Furthermore, we were able to obtain (-)-kopsinine by performing a simultaneous desulfonylation, benzyl hydrogenolysis and diastereoselective alkene reduction of sulfone **28** with Raney nickel<sup>33</sup> to give (-)-kopsinine in only nine steps, which is a significant improvement over the previous 19-step chiral-auxiliary-mediated approach<sup>32</sup>. Our attempts to directly



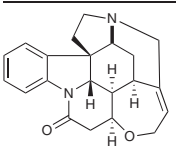
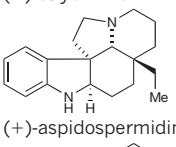
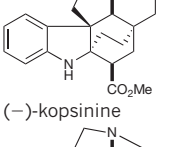
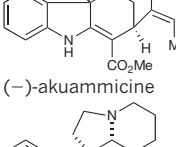
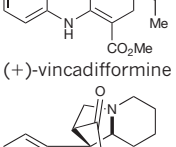
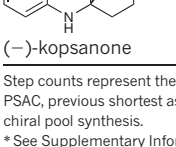
**Figure 7** | Enantioselective total syntheses of (-)-kopsinine and (-)-kopsanone. Reagents and conditions are as follows. **a**, Et<sub>3</sub>N, CH<sub>2</sub>Cl<sub>2</sub>, Me<sub>3</sub>SiI, 0 °C, then MeOH, H<sub>2</sub>C=CHPh<sub>3</sub>Br, 40 °C, then CH<sub>2</sub>Cl<sub>2</sub>, THF, KO<sup>t</sup>-Bu, 0 °C.

**b**, COCl<sub>2</sub>, Et<sub>3</sub>N, toluene, -45 °C to RT, then MeOH, -30 °C to RT. **c**, Pd/C, H<sub>2</sub>, EtOAc, EtOH, 0 °C. **d**, H<sub>2</sub>C=CHSO<sub>2</sub>Ph, benzene, 100 °C. **e**, Raney Ni, EtOH, 78 °C. **f**, 1 N HCl, 130 °C.



**Table 1 | Enantioselective synthesis of six well-known indole alkaloids**


Common intermediate (1)  $\xrightarrow{\text{Synthetic elaboration } \leq 8 \text{ steps}}$  Enantioselective syntheses of *Strychnos*, *Aspidosperma* and *Kopsia* alkaloids

Compound	No. steps here*	Overall yield (%)	PSAC steps	PSCA steps
 (-)-strychnine	12	6.4	25 (refs 19, 20)	16 (ref. 21)
 (+)-aspidospermidine	9	24	13 (ref. 30)	11 (ref. 29)
 (-)-kopsinine	9	14	NA	19 (ref. 32)
 (-)-akuammicine	10	10	NA	NA
 (+)-vincadifformine	11	8.9	NA	10 (ref. 31)
 (-)-kopsanone	11	10	NA	NA

Step counts represent the longest linear sequence from commercially available **9**. NA, not applicable. PSAC, previous shortest asymmetric catalytic synthesis; PSCA, previous shortest chiral auxiliary or chiral pool synthesis.

\* See Supplementary Information for details of these syntheses.

convert (-)-kopsinine to (-)-kopsanone thermally according to the method of ref. 33 were unsuccessful. However, simple acid-mediated hydrolysis to give kopsinic acid (**29**), and subsequent heating of this material without solvent<sup>37</sup>, furnished (-)-kopsanone in only 11 chemical steps.

As anticipated, application of collective total synthesis to each of the target compounds—strychnine, akuammicine, aspidospermidine, vincadifformine, kopsinine and kopsanone—was readily accomplished with unprecedented levels of efficiency (Table 1). Perhaps most notably, these collective asymmetric syntheses took a total of 34 steps for the six natural products described (in comparison with 76 total steps in previous studies).

## Conclusion

We have demonstrated the capabilities of collective total synthesis in combination with organocascade catalysis, a synthetic strategy that provides researchers with the tools to gain ready access to large collections of complex molecular architectures. In particular, we describe the shortest asymmetric synthesis of (-)-strychnine, the best-known

member of the *Strychnos* alkaloid family. We hope to describe the value of this approach in terms of other natural product and medicinal agent families in the near future.

## METHODS SUMMARY

All reactions were performed under an inert atmosphere using dry solvents in anhydrous conditions, unless otherwise noted. Full experimental details and characterization data for all new compounds are included in Supplementary Information.

Received 4 February; accepted 26 May 2011.

- Walji, A. & MacMillan, D. W. C. Strategies to bypass the Taxol problem. Enantioselective cascade catalysis, a new approach for the efficient construction of molecular complexity. *Synlett* 1477–1489 (2007).
- Va, P., Campbell, E. L., Robertson, W. M. & Boger, D. L. Total synthesis and evaluation of a key series of C5-substituted vinblastine derivatives. *J. Am. Chem. Soc.* **132**, 8489–8495 (2010).
- Huang, Y., Walji, A. M., Larsen, C. H. & MacMillan, D. W. C. Enantioselective organocascade catalysis. *J. Am. Chem. Soc.* **127**, 15051–15053 (2005).
- Simmons, B., Walji, A. & MacMillan, D. W. C. Cycle-specific organocascade catalysis: application to olefin hydroamination, hydro-oxidation, and amino-oxidation, and to natural product synthesis. *Angew. Chem. Int. Ed.* **48**, 4349–4353 (2009).
- Dewick, P. M. *Medicinal Natural Products: A Biosynthetic Approach* 3rd edn (Wiley, 2008).
- Corey, E. J., Imai, N. & Pikul, S. Catalytic enantioselective synthesis of a key intermediate for the synthesis of prostanoids. *Tetrahedr. Lett.* **32**, 7517–7520 (1991).
- Kuehne, M. E., Wang, T. & Seraphin, D. The total synthesis of (±)-mossambine. *Synlett* 557–558 (1995).
- Bandarage, U. K., Kuehne, M. E. & Glick, S. D. Total syntheses of racemic albifloranone and its anti-addictive congeners, including 18-methoxycoronaridine. *Tetrahedron* **55**, 9405–9424 (1999).
- Gronal, C., Jeanty, M. & Enders, D. Organocatalytic cascade reactions as a new tool in total synthesis. *Nature Chem.* **2**, 167–178 (2010).
- Bonjoch, J. & Sole, D. Synthesis of strychnine. *Chem. Rev.* **100**, 3455–3482 (2000).
- Sirasani, G., Paul, T., Dougherty, W., Kassel, S. & Andrade, R. B. Concise total syntheses of (±)-strychnine and (±)-akuammicine. *J. Org. Chem.* **75**, 3529–3532 (2010).
- Hudlicky, T. & Reed, J. W. *The Way of Synthesis: Evolution of Design and Methods for Natural Products* (Wiley-VCH, 2007).
- Jones, S. B., Simmons, B. & MacMillan, D. W. C. Nine-step enantioselective total synthesis of (+)-minfiensine. *J. Am. Chem. Soc.* **131**, 13606–13607 (2009).
- Thomas, P. J. & Stirling, C. J. M. Elimination and addition reactions. Part 34. The effect of activating group and medium on leaving group rank in elimination from carbanions. *J. Chem. Soc. Perkin Trans. II* **11**, 1130–1134 (1978).
- Gatta, F. & Misiti, D. Selenium dioxide oxidation of tetrahydro-β-carboline derivatives. *J. Heterocycl. Chem.* **24**, 1183–1187 (1987).
- Prashad, M., Lavecchia, L., Prasad, K. & Repic, O. A convenient synthesis of 3-substituted 1H-indoles. *Synth. Commun.* **25**, 95–100 (1995).
- Oestreich, M. Ed. *The Mizoroki–Heck Reaction* (Wiley, 2009).
- Anet, F. A. L. & Robinson, R. Conversion of the Wieland–Gumlich aldehyde into strychnine. *Chem. Ind.* 245 (1953).
- Knight, S. D. & Overman, L. E. Enantioselective total synthesis of (-)-strychnine. *J. Am. Chem. Soc.* **115**, 9293–9294 (1993).
- Mori, M., Nakanishi, M., Kajishima, D. & Sato, Y. A novel and general synthetic pathway to *Strychnos* indole alkaloids: total syntheses of (-)-tubifoline, (-)-dehydrotubifoline, and (-)-strychnine using palladium-catalyzed asymmetric allylic substitution. *J. Am. Chem. Soc.* **125**, 9801–9807 (2003).
- Sole, D. *et al.* Total synthesis of (-)-strychnine via the Wieland–Gumlich aldehyde. *Angew. Chem. Int. Ed.* **38**, 395–397 (1999).
- Martin, D. B. & Vanderwal, C. D. A synthesis of strychnine by a longest linear sequence of six steps. *Chem. Sci.* **2**, 649–651 (2011).
- Sole, D., Diaba, F. & Bonjoch, J. Nitrogen heterocycles by palladium-catalyzed cyclization of amino-tethered vinyl halides and ketone enolates. *J. Org. Chem.* **68**, 5746–5749 (2003).
- Jeffery, T. On the efficiency of tetraalkylammonium salts in Heck type reactions. *Tetrahedron* **52**, 10113–10130 (1996).
- Marino, J. P., Rubio, M. B., Cao, G. F. & de Dios, A. Total synthesis of (+)-aspidospermidine: a new strategy for the enantiospecific synthesis of *Aspidosperma* alkaloids. *J. Am. Chem. Soc.* **124**, 13398–13399 (2002).
- Kobayashi, S., Peng, G. & Fukuyama, T. Efficient total syntheses of (±)-vincadifformine and (-)-tabersonine. *Tetrahedr. Lett.* **40**, 1519–1522 (1999).
- Hajicek, J. A review on recent developments in syntheses of the post-second indole alkaloids. Part I: The primary alkaloid types. *Collect. Czech. Chem. Commun.* **69**, 1681–1767 (2004).
- Cho, H.-K., Tam, N. T. & Cho, C. G. Total synthesis of (±) aspidospermidine starting from 3-ethyl-5-bromo-2-pyrone. *Bull. Korean Chem. Soc.* **31**, 3382–3384 (2010).
- Gnecco, D. *et al.* Synthesis of an aspidosperma alkaloid precursor: synthesis of (+)-aspidospermidine. *Arkivoc* **2003** (xi), 185–192 (2003).
- Kozmin, S. A., Iwama, T., Huang, Y. & Rawal, V. H. An efficient approach to *Aspidosperma* alkaloids via [4 + 2] cycloadditions of aminosiloxycyclopropanes: Stereocontrolled total synthesis of (±)-tabersonine. Gram-scale catalytic

- asymmetric syntheses of (+)-tabersonine and (+)-16-methoxytabersonine. Asymmetric syntheses of (+)-aspidospermidine and (–)-quebrachamine. *J. Am. Chem. Soc.* **124**, 4628–4641 (2002).
31. Kuehne, M. *et al.* Application of ferrocenylalkyl chiral auxiliaries to syntheses of indolenine alkaloids: enantioselective syntheses of vincadifformine,  $\psi$ - and 20-*epi*- $\psi$ -vincadifformines, tabersonine, ibophyllidine, and mossambine. *J. Org. Chem.* **63**, 2172–2183 (1998).
  32. Magnus, P. & Brown, P. Total synthesis of (–)-kopsinilam, (–)-kopsinine, and the bis-indole alkaloids (–)-norpleiomutine and (–)-pleiomutine. *J. Chem. Soc. Chem. Commun.* 184–186 (1985).
  33. Kuehne, M. E. & Seaton, P. J. Studies in biomimetic alkaloid syntheses. 13. Total syntheses of racemic aspidofractine, pleiocarpine, pleiocarpinine, kopsinine, *N*-methylkopsanone, and kopsanone. *J. Org. Chem.* **50**, 4790–4796 (1985).
  34. Wenkert, E. & Pestchanker, M. J. A formal total synthesis of kopsinine. *J. Org. Chem.* **53**, 4875–4877 (1988).
  35. Ogawa, M., Kitagawa, Y. & Natsume, M. A high-yield cyclization reaction for the framework of aspidosperma alkaloids synthesis of (±)-kopsinine and its related alkaloids. *Tetrahedr. Lett.* **28**, 3985–3986 (1987).
  36. Gallagher, T. & Magnus, P. Synthesis of (±)-kopsanone and (±)-10,22-dioxokopsane, heptacyclic indole alkaloids. *J. Am. Chem. Soc.* **105**, 2086–2087 (1983).
  37. Kump, C., Dugan, J. J. & Schmid, H. Ringschlussreaktionen an Pleiocarpa-Alkaloiden. *Helv. Chim. Acta* **49**, 1237–1243 (1966).
  38. Magnus, P., Payne, A. H. & Hobson, L. Synthesis of the kopsia alkaloids (±)-11,12-demethoxylahadinine B, (±)-kopsidasine and (±)-kopsidasine-*N*-oxide. *Tetrahedr. Lett.* **41**, 2077–2081 (2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Financial support was provided by NIHGMMS (R01 GM078201-05) and gifts from Merck, Bristol-Myers Squibb and Abbott. S.B.J. and B.S. thank Bristol-Myers Squibb and Merck, respectively, for graduate fellowships.

**Author Contributions** S.B.J., B.S. and A.M. participated in the performance and analysis of the experiments. S.B.J., B.S., A.M. and D.W.C.M. designed the experiments. S.B.J. and D.W.C.M. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.W.C.M. ([dmacmill@princeton.edu](mailto:dmacmill@princeton.edu)).



# Genome sequence and analysis of the tuber crop potato

The Potato Genome Sequencing Consortium\*

Potato (*Solanum tuberosum* L.) is the world's most important non-grain food crop and is central to global food security. It is clonally propagated, highly heterozygous, autotetraploid, and suffers acute inbreeding depression. Here we use a homozygous doubled-monoploid potato clone to sequence and assemble 86% of the 844-megabase genome. We predict 39,031 protein-coding genes and present evidence for at least two genome duplication events indicative of a palaeopolyploid origin. As the first genome sequence of an asterid, the potato genome reveals 2,642 genes specific to this large angiosperm clade. We also sequenced a heterozygous diploid clone and show that gene presence/absence variants and other potentially deleterious mutations occur frequently and are a likely cause of inbreeding depression. Gene family expansion, tissue-specific expression and recruitment of genes to new pathways contributed to the evolution of tuber development. The potato genome sequence provides a platform for genetic improvement of this vital crop.

Potato (*Solanum tuberosum* L.) is a member of the Solanaceae, an economically important family that includes tomato, pepper, aubergine (eggplant), petunia and tobacco. Potato belongs to the asterid clade of eudicot plants that represents ~25% of flowering plant species and from which a complete genome sequence has not yet, to our knowledge, been published. Potato occupies a wide eco-geographical range<sup>1</sup> and is unique among the major world food crops in producing stolons (underground stems) that under suitable environmental conditions swell to form tubers. Its worldwide importance, especially within the developing world, is growing rapidly, with production in 2009 reaching 330 million tons (<http://www.fao.org>). The tubers are a globally important dietary source of starch, protein, antioxidants and vitamins<sup>2</sup>, serving the plant as both a storage organ and a vegetative propagation system. Despite the importance of tubers, the evolutionary and developmental mechanisms of their initiation and growth remain elusive.

Outside of its natural range in South America, the cultivated potato is considered to have a narrow genetic base resulting originally from limited germplasm introductions to Europe. Most potato cultivars are autotetraploid ( $2n = 4x = 48$ ), highly heterozygous, suffer acute inbreeding depression, and are susceptible to many devastating pests and pathogens, as exemplified by the Irish potato famine in the mid-nineteenth century. Together, these attributes present a significant barrier to potato improvement using classical breeding approaches. A challenge to the scientific community is to obtain a genome sequence that will ultimately facilitate advances in breeding.

To overcome the key issue of heterozygosity and allow us to generate a high-quality draft potato genome sequence, we used a unique homozygous form of potato called a doubled monoploid, derived using classical tissue culture techniques<sup>3</sup>. The draft genome sequence from this genotype, *S. tuberosum* group Phureja DM1-3 516 R44 (hereafter referred to as DM), was used to integrate sequence data from a heterozygous diploid breeding line, *S. tuberosum* group Tuberosum RH89-039-16 (hereafter referred to as RH). These two genotypes represent a sample of potato genomic diversity; DM with its fingerling (elongated) tubers was derived from a primitive South American cultivar whereas RH more closely resembles commercially cultivated tetraploid potato. The combined data resources, allied to

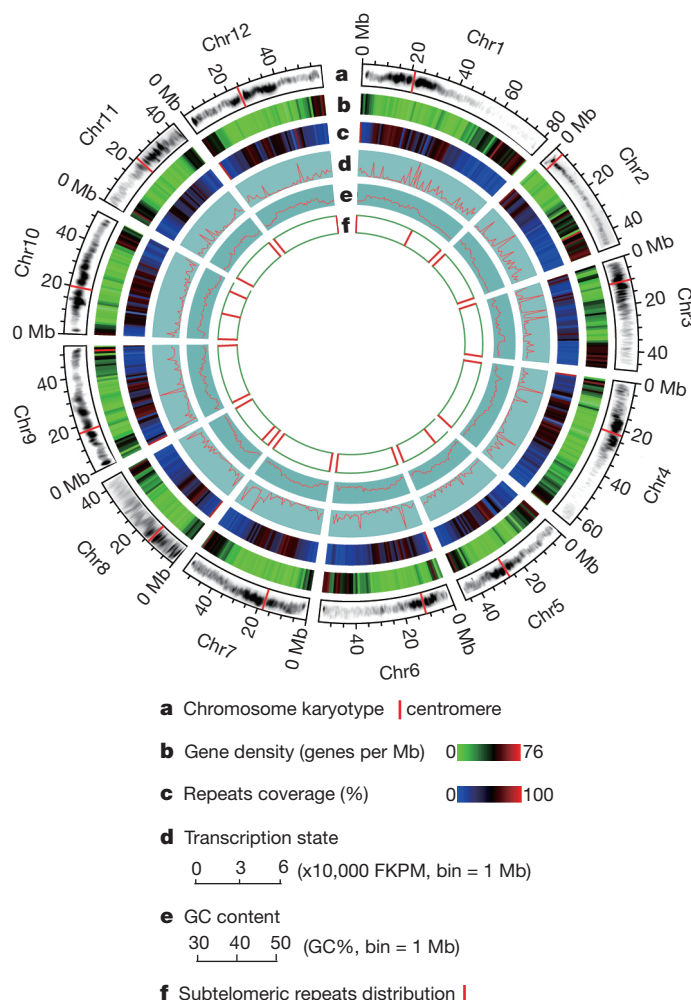
deep transcriptome sequence from both genotypes, allowed us to explore potato genome structure and organization, as well as key aspects of the biology and evolution of this important crop.

## Genome assembly and annotation

We sequenced the nuclear and organellar genomes of DM using a whole-genome shotgun sequencing (WGS) approach. We generated 96.6 Gb of raw sequence from two next-generation sequencing (NGS) platforms, Illumina Genome Analyser and Roche Pyrosequencing, as well as conventional Sanger sequencing technologies. The genome was assembled using SOAPdenovo<sup>4</sup>, resulting in a final assembly of 727 Mb, of which 93.9% is non-gapped sequence. Ninety per cent of the assembly falls into 443 superscaffolds larger than 349 kb. The 17-nucleotide depth distribution (Supplementary Fig. 1) suggests a genome size of 844 Mb, consistent with estimates from flow cytometry<sup>5</sup>. Our assembly of 727 Mb is 117 Mb less than the estimated genome size. Analysis of the DM scaffolds indicates 62.2% repetitive content in the assembled section of the DM genome, less than the 74.8% estimated from bacterial artificial chromosome (BAC) and fosmid end sequences (Supplementary Table 1), indicating that much of the unassembled genome is composed of repetitive sequences.

We assessed the quality of the WGS assembly through alignment to Sanger-derived phase 2 BAC sequences. In an alignment length of ~1 Mb (99.4% coverage), no gross assembly errors were detected (Supplementary Table 2 and Supplementary Fig. 2). Alignment of fosmid and BAC paired-end sequences to the WGS scaffolds revealed limited ( $\leq 0.12\%$ ) potential misassemblies (Supplementary Table 3). Extensive coverage of the potato genome in this assembly was confirmed using available expressed sequence tag (EST) data; 97.1% of 181,558 available Sanger-sequenced *S. tuberosum* ESTs (>200 bp) were detected. Repetitive sequences account for at least 62.2% of the assembled genome (452.5 Mb) (Supplementary Table 1) with long terminal repeat retrotransposons comprising the majority of the transposable element classes, representing 29.4% of the genome. In addition, subtelomeric repeats were identified at or near chromosomal ends (Fig. 1). Using a newly constructed genetic map based on 2,603 polymorphic markers in conjunction with other available

\*Lists of authors and their affiliations appear at the end of the paper.



**Figure 1 | The potato genome.** **a**, Ideograms of the 12 pseudochromosomes of potato (in Mb scales). Each of the 12 pachytene chromosomes from DM was digitally aligned with the ideogram (the amount of DNA in each unit of the pachytene chromosomes is not in proportion to the scales of the pseudochromosomes). **b**, Gene density represented as number of genes per Mb (non-overlapping, window size = 1 Mb). **c**, Percentage of coverage of repetitive sequences (non-overlapping windows, window size = 1 Mb). **d**, Transcription state. The transcription level for each gene was estimated by averaging the fragments per kb exon model per million mapped reads (FPKM) from different tissues in non-overlapping 1-Mb windows. **e**, GC content was estimated by the per cent G+C in 1-Mb non-overlapping windows. **f**, Distribution of the subtelomeric repeat sequence CL14\_cons.

genetic and physical maps, we genetically anchored 623 Mb (86%) of the assembled genome (Supplementary Fig. 3), and constructed pseudomolecules for each of the 12 chromosomes (Fig. 1), which harbour 90.3% of the predicted genes.

To aid annotation and address a series of biological questions, we generated 31.5 Gb of RNA-Seq data from 32 DM and 16 RH libraries representing all major tissue types, developmental stages and responses to abiotic and biotic stresses (Supplementary Table 4). For annotation, reads were mapped against the DM genome sequence (90.2% of 824,621,408 DM reads and 88.6% of 140,375,647 RH reads) and in combination with *ab initio* gene prediction, protein and EST alignments, we annotated 39,031 protein-coding genes. RNA-Seq data revealed alternative splicing; 9,875 genes (25.3%) encoded two or more isoforms, indicative of more functional variation than represented by the gene set alone. Overall, 87.9% of the gene models were supported by transcript and/or protein similarity with only 12.1% derived solely from *ab initio* gene predictions (Supplementary Table 5).

Karyotypes of RH and DM suggested similar heterochromatin content<sup>6</sup> (Supplementary Table 6 and Supplementary Fig. 4) with large blocks of heterochromatin located at the pericentromeric regions (Fig. 1). As observed in other plant genomes, there was an inverse relationship between gene density and repetitive sequences (Fig. 1). However, many predicted genes in heterochromatic regions are expressed, consistent with observations in tomato<sup>7</sup> that genic 'islands' are present in the heterochromatic 'ocean'.

## Genome evolution

Potato is the first sequenced genome of an asterid, a clade within eudicots that encompasses nearly 70,000 species characterized by unique morphological, developmental and compositional features<sup>8</sup>. Orthologous clustering of the predicted potato proteome with 11 other green plant genomes revealed 4,479 potato genes in 3,181 families in common (Fig. 2a); 24,051 potato genes clustered with at least one of the 11 genomes. Filtering against transposable elements and 153 nonasterid and 57 asterid publicly available transcript-sequence data sets yielded 2,642 high-confidence asterid-specific and 3,372 potato-lineage-specific genes (Supplementary Fig. 5); both sets were enriched for genes of unknown function that had less expression support than the core Viridiplantae genes. Genes encoding transcription factors, self-incompatibility, and defence-related proteins were evident in the asterid-specific gene set (Supplementary Table 7) and presumably contribute to the unique characteristics of asterids.

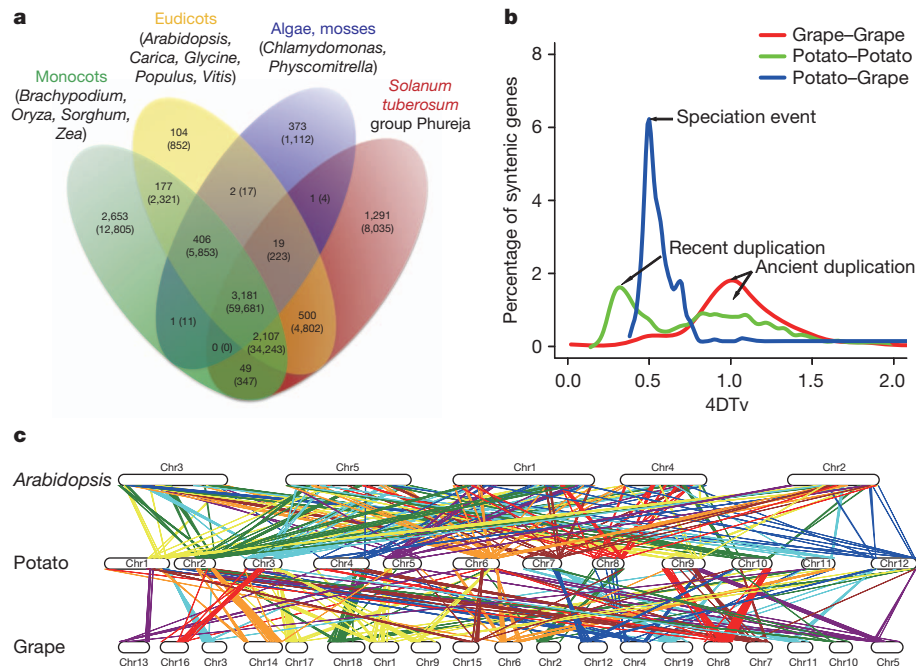
Structurally, we identified 1,811 syntenic gene blocks involving 10,046 genes in the potato genome (Supplementary Table 8). On the basis of these pairwise paralogous segments, we calculated an age distribution based on the number of transversions at fourfold degenerate sites (4DTV) for all duplicate pairs. In general, two significant groups of blocks are seen in the potato genome (4DTV ~0.36 and ~1.0; Fig. 2b), suggesting two whole-genome duplication (WGD) events. We also identified collinear blocks between potato and three rosoid genomes (*Vitis vinifera*, *Arabidopsis thaliana* and *Populus trichocarpa*) that also suggest both events (Fig. 2c and Supplementary Fig. 6). The ancient WGD corresponds to the ancestral hexaploidization ( $\gamma$ ) event in grape (Fig. 2b), consistent with a previous report based on EST analysis that the two main branches of eudicots, the asterids and rosids, may share the same palaeo-hexaploid duplication event<sup>9</sup>. The  $\gamma$  event probably occurred after the divergence between dicots and monocots about  $185 \pm 55$  million years ago<sup>10</sup>. The recent duplication can therefore be placed at ~67 million years ago, consistent with the WGD that occurred near the Cretaceous-Tertiary boundary (~65 million years ago)<sup>11</sup>. The divergence of potato and grape occurred at ~89 million years ago (4DTV ~0.48), which is likely to represent the split between the rosids and asterids.

## Haplotype diversity

High heterozygosity and inbreeding depression are inherent to potato, a species that predominantly outcrosses and propagates by means of vegetative organs. Indeed, the phenotypes of DM and RH differ, with RH more vigorous than DM (Fig. 3a). To explore the extent of haplotype diversity and possible causes of inbreeding depression, we sequenced and assembled 1,644 RH BAC clones generating 178 Mb of non-redundant sequence from both haplotypes (~10% of the RH genome with uneven coverage) (Supplementary Tables 9–11). After filtering to remove repetitive sequences, we aligned 99 Mb of RH sequence (55%) to the DM genome. These regions were largely collinear with an overall sequence identity of 97.5%, corresponding to one single-nucleotide polymorphism (SNP) every 40 bp and one insertion/deletion (indel) every 394 bp (average length 12.8 bp). Between the two RH haplotypes, 6.6 Mb of sequence could be aligned with 96.5% identity, corresponding to 1 SNP per 29 bp and 1 indel per 253 bp (average length 10.4 bp).

Current algorithms are of limited use in *de novo* whole-genome assembly or haplotype reconstruction of highly heterozygous genomes

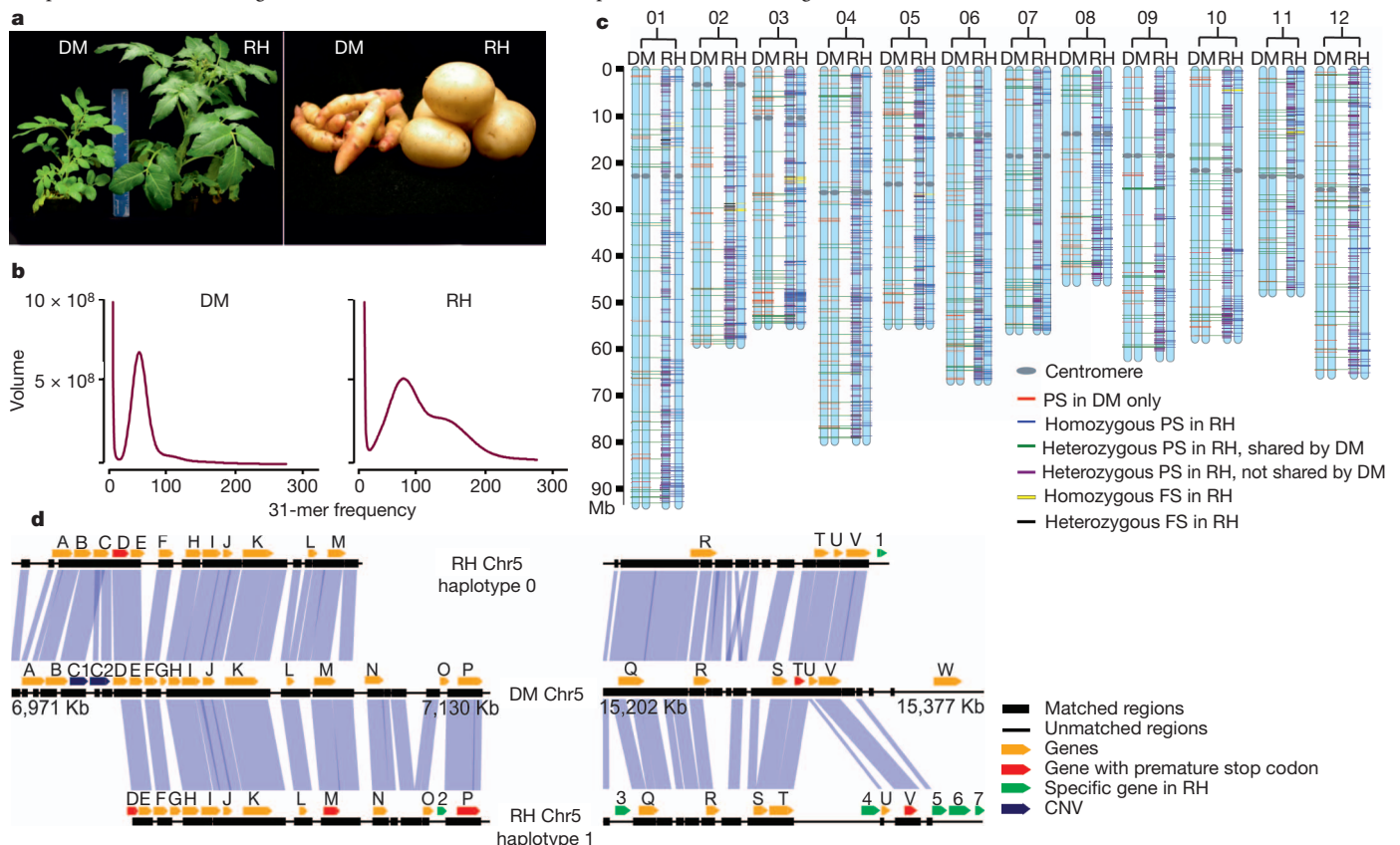




**Figure 2 | Comparative analyses and evolution of the potato genome.**

**a**, Clusters of orthologous and paralogous gene families in 12 plant species as identified by OrthoMCL<sup>33</sup>. Gene family number is listed in each of the components; the number of genes within the families for all of the species

within the component is noted within parentheses. **b**, Genome duplication in dicot genomes as revealed through 4Dtv analyses. **c**, Syntenic blocks between *A. thaliana*, potato, and *V. vinifera* (grape) demonstrating a high degree of conserved gene order between these taxa.



**Figure 3 | Haplotype diversity and inbreeding depression.** **a**, Plants and tubers of DM and RH showing that RH has greater vigour. **b**, Illumina K-mer volume histograms of DM and RH. The volume of K-mers (y-axis) is plotted against the frequency at which they occur (x-axis). The leftmost truncated peaks at low frequency and high volume represent K-mers containing essentially random sequencing errors, whereas the distribution to the right represents proper (putatively error-free) data. In contrast to the single modality of DM, RH exhibits clear bi-modality caused by heterozygosity. **c**, Genomic distribution of premature

stop, frameshift and presence/absence variation mutations contributing to inbreeding depression. The hypothetical RH pseudomolecules were solely inferred from the corresponding DM ones. Owing to the inability to assign heterozygous PS and FS of RH to a definite haplotype, all heterozygous PS and FS were arbitrarily mapped to the left haplotype of RH. **d**, A zoom-in comparative view of the DM and RH genomes. The left and right alignments are derived from the euchromatic and heterochromatic regions of chromosome 5, respectively. Most of the gene annotations, including PS and RH-specific genes, are supported by transcript data.

such as RH, as shown by K-mer frequency count histograms (Fig. 3b and Supplementary Table 12). To complement the BAC-level comparative analysis and provide a genome-wide perspective of heterozygosity in RH, we mapped 1,118 million whole-genome NGS reads from RH (84× coverage) onto the DM assembly. A total of 457.3 million reads uniquely aligned providing 90.6% (659.1 Mb) coverage. We identified 3.67 million SNPs between DM and one or both haplotypes of RH, with an error rate of 0.91% based on evaluation of RH BAC sequences. We used this data set to explore the possible causes of inbreeding depression by quantifying the occurrence of premature stop, frameshift and presence/absence variants<sup>12</sup>, as these disable gene function and contribute to genetic load (Supplementary Tables 13–16). We identified 3,018 SNPs predicted to induce premature stop codons in RH, with 606 homozygous (in both haplotypes) and 2,412 heterozygous. In DM, 940 premature stop codons were identified. In the 2,412 heterozygous RH premature stop codons, 652 were shared with DM and the remaining 1,760 were found in RH only (Fig. 3c and Supplementary Table 13). Frameshift mutations were identified in 80 loci within RH, 49 homozygous and 31 heterozygous, concentrated in seven genomic regions (Fig. 3c and Supplementary Table 14). Finally, we identified presence/absence variations for 275 genes; 246 were RH specific (absent in DM) and 29 were DM specific, with 125 and 9 supported by RNA-Seq and/or Gene Ontology<sup>13</sup> annotation for RH and DM, respectively (Supplementary Tables 15 and 16). Collectively, these data indicate that the complement of homozygous deleterious alleles in DM may be responsible for its reduced level of vigour (Fig. 3a).

The divergence between potato haplotypes is similar to that reported between out-crossing maize accessions<sup>14</sup> and, coupled with our inability to successfully align 45% of the BAC sequences, intra- and inter-genome diversity seem to be a significant feature of the potato genome. A detailed comparison of the three haplotypes (DM and the two haplotypes of RH) at two genomic regions (334 kb in length) using the RH BAC sequence (Fig. 3d and Supplementary Tables 17 and 18) revealed considerable sequence and structural variation. In one region ('euchromatic'; Fig. 3d) we observed one instance of copy number variation, five genes with premature stop codons, and seven RH-specific genes. These observations indicate that the plasticity of the potato genome is greater than revealed from the unassembled RH NGS. Improved assembly algorithms, increased read lengths, and *de novo* sequences of additional haplotypes will reveal the full catalogue of genes critical to inbreeding depression.

## Tuber biology

In developing DM and RH tubers, 15,235 genes were expressed in the transition from stolons to tubers, with 1,217 transcripts exhibiting >5-fold expression in stolons versus five RH tuber tissues (young tuber, mature tuber, tuber peel, cortex and pith; Supplementary Table 19). Of these, 333 transcripts were upregulated during the transition from stolon to tuber, with the most highly upregulated transcripts encoding storage proteins. Foremost among these were the genes encoding proteinase inhibitors and patatin (15 genes), in which the phospholipase A function has been largely replaced by a protein storage function in the tuber<sup>15</sup>. In particular, a large family of 28 Kunitz protease inhibitor genes (KTIs) was identified with twice the number of genes in potato compared to tomato. The KTI genes are distributed across the genome with individual members exhibiting specific expression patterns (Fig. 4a, b). KTIs are frequently induced after pest and pathogen attack and act primarily as inhibitors of exogenous proteinases<sup>16</sup>; therefore the expansion of the KTI family may provide resistance to biotic stress for the newly evolved vulnerable underground organ.

The stolon to tuber transition also coincides with strong upregulation of genes associated with starch biosynthesis (Fig. 4c). We observed several starch biosynthetic genes that were 3–8-fold more highly expressed in tuber tissues of RH compared to DM (Fig. 4c). Together this suggests a stronger shift from the relatively low sink strength of the ATP-generating general carbon metabolism reactions

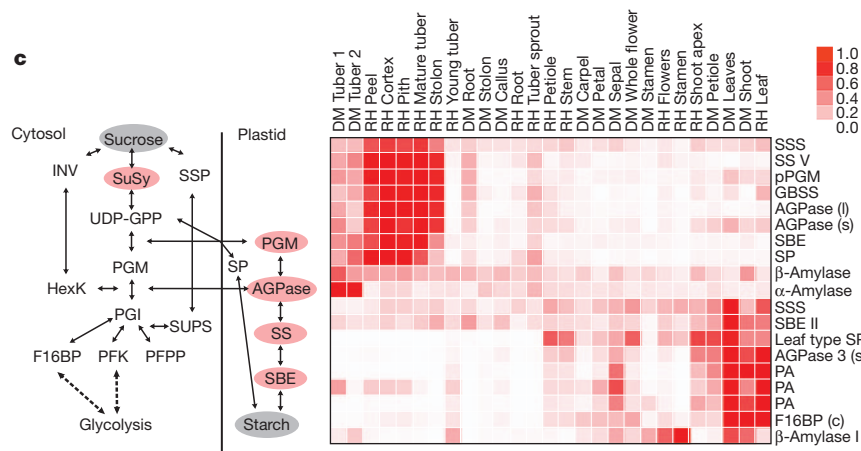
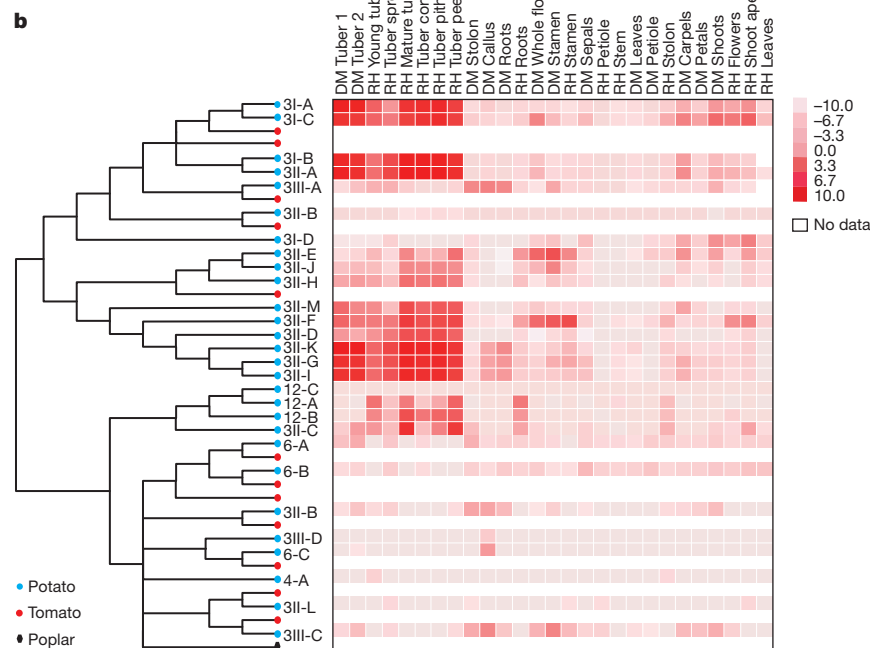
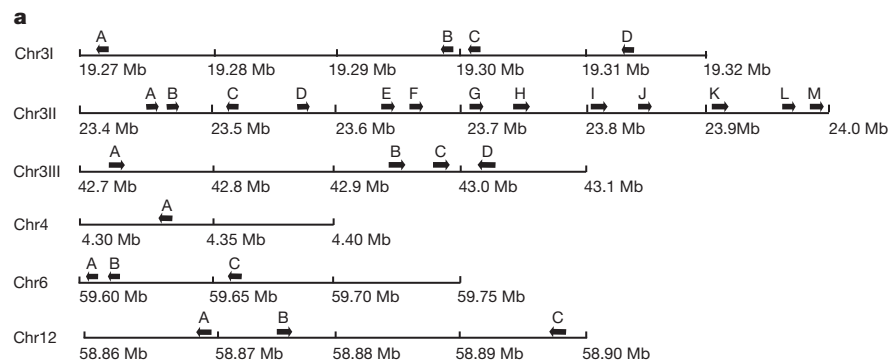
towards the plastidic starch synthesis pathway in tubers of RH, thereby causing a flux of carbon into the amyloplast. This contrasts with the cereal endosperm where carbon is transported into the amyloplast in the form of ADP-glucose via a specific transporter (brittle 1 protein<sup>17</sup>). Carbon transport into the amyloplasts of potato tubers is primarily in the form of glucose-6-phosphate<sup>18</sup>, although recent evidence indicates that glucose-1-phosphate is quantitatively important under certain conditions<sup>19</sup>. The transport mechanism for glucose-1-phosphate is unknown and the genome sequence contains six genes for hexose-phosphate transporters with two highly and specifically expressed in stolons and tubers. Furthermore, an additional 23 genes encode proteins homologous to other carbohydrate derivative transporters, such as triose phosphate, phosphoenolpyruvate, or UDP-glucuronic acid transporters and two loci with homologues for the brittle 1 protein. By contrast, in leaves, carbon-fixation-specific genes such as plastidic aldolase, fructose-1,6-biphosphatase and distinct leaf isoforms of starch synthase, starch branching enzyme, starch phosphorylase and ADP-glucose pyrophosphorylase were upregulated. Of particular interest is the difference in tuber expression of enzymes involved in the hydrolytic and phosphorolytic starch degradation pathways. Considerably greater levels of  $\alpha$ -amylase (10–25-fold) and  $\beta$ -amylase (5–10-fold) mRNAs were found in DM tubers compared to RH, whereas  $\alpha$ -1,4 glucan phosphorylase mRNA was equivalent in DM and RH tubers. These gene expression differences between the breeding line RH and the more primitive DM are consistent with the concept that increasing tuber yield may be partially attained by selection for decreased activity of the hydrolytic starch degradation pathway.

Recent studies using a potato genotype strictly dependent on short days for tuber induction (*S. tuberosum* group Andigena) identified a potato homologue (*SP6A*) of *A. thaliana* *FLOWERING LOCUS T* (*FT*) as the long-distance tuberization inductive signal. *SP6A* is produced in the leaves, consistent with its role as the mobile signal (S. Prat, personal communication). *SP/FT* is a multi-gene family (Supplementary Text and Supplementary Fig. 7) and expression of a second *FT* homologue, *SP5G*, in mature tubers suggests a possible function in the control of tuber sprouting, a photoperiod-dependent phenomenon<sup>20</sup>. Likewise, expression of a homologue of the *A. thaliana* flowering time MADS box gene *SOC1*, acting downstream of *FT*<sup>21</sup>, is restricted to tuber sprouts (Supplementary Fig. 8). Expression of a third *FT* homologue, *SP3D*, does not correlate with tuberization induction but instead with transition to flowering, which is regulated independently of day length (S. Prat, personal communication). These data indicate that neofunctionalization of the day-length-dependent flowering control pathway has occurred in potato to control formation and possibly sprouting of a novel storage organ, the tuber (Supplementary Fig. 9).

## Disease resistance

Potato is susceptible to a wide range of pests and pathogens and the identification of genes conferring disease resistance has been a major focus of the research community. Most cloned disease resistance genes in the Solanaceae encode nucleotide-binding site (NBS) and leucine-rich-repeat (LRR) domains. The DM assembly contains 408 NBS-LRR-encoding genes, 57 Toll/interleukin-1 receptor/plant R gene homology (TIR) domains and 351 non-TIR types (Supplementary Table 20), similar to the 402 resistance (*R*) gene candidates in *Populus*<sup>22</sup>. Highly related homologues of the cloned potato late blight resistance genes *R1*, *RB*, *R2*, *R3a*, *Rpi-blb2* and *Rpi-vnt1.1* were present in the assembly. In RH, the chromosome 5 *R1* cluster contains two distinct haplotypes; one is collinear with the *R1* region in DM (Supplementary Fig. 10), yet neither the DM nor the RH *R1* regions are collinear with other potato *R1* regions<sup>23,24</sup>. Comparison of the DM potato *R* gene sequences with well-established gene models (functional *R* genes) indicates that many NBS-LRR genes (39.4%) are pseudogenes owing to indels, frameshift mutations, or premature stop





**Figure 4 | Gene expression of selected tissues and genes.** **a**, KTI gene organization across the potato genome. Black arrows indicate the location of individual genes on six scaffolds located on four chromosomes. **b**, Phylogenetic tree and KTI gene expression heat map. The KTI genes were clustered using all potato and tomato genes available with the *Populus* KTI gene as an out-group. The tissue specificity of individual members of the highly expanded potato gene family is shown in the heat map. Expression levels are indicated by shades of red, where white indicates no expression or lack of data for tomato and poplar. **c**, A model of starch synthesis showing enzyme activities is shown on the left. AGPase, ADP-glucose pyrophosphorylase; F16BP, fructose-1,6-biphosphatase; HexK, hexokinase; INV, invertase; PFK, phosphofructokinase; PFPP, pyrophosphate-fructose-6-phosphate-1-phosphotransferase; PGI, phosphoglucose isomerase; PGM, phosphoglucomutase; SBE, starch branching enzyme; SP, starch phosphorylase; SPP, sucrose phosphate phosphatase; SS, starch synthase; SuSy, sucrose synthase; SUPS, sucrose phosphate synthase; UDP-GPP, UDP-glucose pyrophosphorylase. The grey background denotes substrate (sucrose) and product (starch) and the red background indicates genes that are specifically upregulated in RH versus DM. On the right, a heat map of the genes involved in carbohydrate metabolism is shown. ADP-glucose pyrophosphorylase large subunit, AGPase (l); ADP-glucose pyrophosphorylase small subunit, AGPase (s); ADP-glucose pyrophosphorylase small subunit 3, AGPase 3 (s); cytosolic fructose-1,6-biphosphatase, F16BP (c); granule bound starch synthase, GBSS; leaf type L starch phosphorylase, Leaf type SP; plastidic phosphoglucomutase, pPGM; starch branching enzyme II, SBE II; soluble starch synthase, SSS; starch synthase V, SSV; three variants of plastidic aldolase, PA.

codons including the *R1*, *R3a* and *Rpi-vnt1.1* clusters that contain extensive chimaeras and exhibit evolutionary patterns of type I *R* genes<sup>25</sup>. This high rate of pseudogenization parallels the rapid evolution of effector genes observed in the potato late blight pathogen, *Phytophthora infestans*<sup>26</sup>. Coupled with abundant haplotype diversity, tetraploid potato may therefore contain thousands of *R*-gene analogues.

## Conclusions and future directions

We sequenced a unique doubled-monoploid potato clone to overcome the problems associated with genome assembly due to high levels of

heterozygosity and were able to generate a high-quality draft potato genome sequence that provides new insights into eudicot genome evolution. Using a combination of data from the vigorous, heterozygous diploid RH and relatively weak, doubled-monoploid DM, we could directly address the form and extent of heterozygosity in potato and provide the first view into the complexities that underlie inbreeding depression. Combined with other recent studies, the potato genome sequence may elucidate the evolution of tuberization. This evolutionary innovation evolved exclusively in the *Solanum* section *Petota* that encompasses ~200 species distributed from the southwestern United States to central Argentina and Chile. Neighbouring *Solanum* species,

including the *Lycopersicon* section, which comprises wild and cultivated tomatoes, did not acquire this trait. Both gene family expansion and recruitment of existing genes for new pathways contributed to the evolution of tuber development in potato.

Given the pivotal role of potato in world food production and security, the potato genome provides a new resource for use in breeding. Many traits of interest to plant breeders are quantitative in nature and the genome sequence will simplify both their characterization and deployment in cultivars. Whereas much genetic research is conducted at the diploid level in potato, almost all potato cultivars are tetraploid and most breeding is conducted in tetraploid material. Hence, the development of experimental and computational methods for routine and informative high-resolution genetic characterization of polyploids remains an important goal for the realization of many of the potential benefits of the potato genome sequence.

## METHODS SUMMARY

DM1-3 516 R44 (DM) resulted from chromosome doubling of a monoploid ( $1n = 1x = 12$ ) derived by anther culture of a heterozygous diploid ( $2n = 2x = 24$ ) *S. tuberosum* group Phureja clone (PI 225669)<sup>27</sup>. RH89-039-16 (RH) is a diploid clone derived from a cross between a *S. tuberosum* 'dihaploid' (SUH2293) and a diploid clone (BC1034) generated from a cross between two *S. tuberosum* × *S. tuberosum* group Phureja hybrids<sup>28</sup> (Supplementary Fig. 11). Sequence data from three platforms, Sanger, Roche 454 Pyrosequencing, and Illumina Sequencing-by-Synthesis, were used to assemble the DM genome using the SOAPdenovo assembly algorithm<sup>4</sup>. The RH genotype was sequenced using shotgun sequencing of BACs and WGS in which reads were mapped to the DM reference assembly. Superscaffolds were anchored to the 12 linkage groups using a combination of *in silico* and genetic mapping data. Repeat sequences were identified through sequence similarity at the nucleotide and protein level<sup>29</sup>. Genes were annotated using a combined approach<sup>30</sup> on the repeat masked genome with *ab initio* gene predictions, protein similarity and transcripts to build optimal gene models. Illumina RNA-Seq reads were mapped to the DM draft sequence using Tophat<sup>31</sup> and expression levels from the representative transcript were determined using Cufflinks<sup>32</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 11 January; accepted 3 May 2011.

Published online 10 July 2011.

- Hijmans, R. J. Global distribution of the potato crop. *Am. J. Potato Res.* **78**, 403–412 (2001).
- Burlingame, B., Mouillé, B. & Charrondière, R. Nutrients, bioactive non-nutrients and anti-nutrients in potatoes. *J. Food Compos. Anal.* **22**, 494–502 (2009).
- Paz, M. M. & Veilleux, R. E. Influence of culture medium and *in vitro* conditions on shoot regeneration in *Solanum phureja* monoploids and fertility of regenerated doubled monoploids. *Plant Breed.* **118**, 53–57 (1999).
- Li, R. *et al.* De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272 (2010).
- Arumuganathan, K. & Earle, E. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **9**, 208–218 (1991).
- Tang, X. *et al.* Assignment of genetic linkage maps to diploid *Solanum tuberosum* pachytene chromosomes by BAC-FISH technology. *Chromosome Res.* **17**, 899–915 (2009).
- Peters, S. A. *et al.* *Solanum lycopersicon* cv. Heinz 1706 chromosome 6: distribution and abundance of genes and retrotransposable elements. *Plant J.* **58**, 857–869 (2009).
- Albach, D. C., Soltis, P. S. & Soltis, D. E. Patterns of embryological and biochemical evolution in the Asterids. *Syst. Bot.* **26**, 242–262 (2001).
- Tang, H. *et al.* Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* **18**, 1944–1954 (2008).
- Jaillon, O. *et al.* The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463–467 (2007).
- Fawcett, J. A., Maere, S. & Van de Peer, Y. Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *Proc. Natl Acad. Sci. USA* **106**, 5737–5742 (2009).
- Lai, J. *et al.* Genome-wide patterns of genetic variation among elite maize inbred lines. *Nature Genet.* **42**, 1027–1030 (2010).
- Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature Genet.* **25**, 25–29 (2000).
- Gore, M. A. *et al.* A first-generation haplotype map of maize. *Science* **326**, 1115–1117 (2009).
- Prat, S. *et al.* Gene expression during tuber development in potato plants. *FEBS Lett.* **268**, 334–338 (1990).
- Glaczinski, H., Heibges, A., Salamini, R. & Gebhardt, C. Members of the Kunitz-type protease inhibitor gene family of potato inhibit soluble tuber invertase *in vitro*. *Potato Res.* **45**, 163–176 (2002).
- Shannon, J. C., Pien, F. M. & Liu, K. C. Nucleotides and nucleotide sugars in developing maize endosperms: synthesis of ADP-glucose in *brittle-1*. *Plant Physiol.* **110**, 835–843 (1996).
- Tauberger, E. *et al.* Antisense inhibition of plastidial phosphoglucomutase provides compelling evidence that potato tuber amyloplasts import carbon from the cytosol in the form of glucose-6-phosphate. *Plant J.* **23**, 43–53 (2000).
- Fettke, J. *et al.* Glucose 1-phosphate is efficiently taken up by potato (*Solanum tuberosum*) tuber parenchyma cells and converted to reserve starch granules. *New Phytol.* **185**, 663–675 (2010).
- Sonnenwald, U. Control of potato tuber sprouting. *Trends Plant Sci.* **6**, 333–335 (2001).
- Yoo, S. K. *et al.* CONSTANS activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through FLOWERING LOCUS T to promote flowering in *Arabidopsis*. *Plant Physiol.* **139**, 770–778 (2005).
- Kohler, A. *et al.* Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant Mol. Biol.* **66**, 619–636 (2008).
- Ballvora, A. *et al.* Comparative sequence analysis of *Solanum* and *Arabidopsis* in a hot spot for pathogen resistance on potato chromosome V reveals a patchwork of conserved and rapidly evolving genome segments. *BMC Genomics* **8**, 112 (2007).
- Kuang, H. *et al.* The R1 resistance gene cluster contains three groups of independently evolving, type I R1 homologues and shows substantial structural variation among haplotypes of *Solanum demissum*. *Plant J.* **44**, 37–51 (2005).
- Kuang, H., Woo, S. S., Meyers, B. C., Nevo, E. & Michelmore, R. W. Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* **16**, 2870–2894 (2004).
- Haas, B. J. *et al.* Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **461**, 393–398 (2009).
- Haynes, F. L. In *Prospects for the Potato in the Developing World: an International Symposium on Key Problems and Potentials for Greater Use of the Potato in the Developing World* (ed. French, E. R.) 100–110 (International Potato Center (CIP), 1972).
- van Os, H. *et al.* Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. *Genetics* **173**, 1075–1087 (2006).
- Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **25**, 4.10.1–4.10.14 (2004).
- Elsik, C. G. *et al.* Creating a honey bee consensus gene set. *Genome Biol.* **8**, R13 (2007).
- Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
- Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28**, 511–515 (2010).
- Li, L., Stoeckert, C. J. Jr & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge the assistance of W. Amorós, B. Babinska, R. V. Baslerov, B. K. Bumazhkin, M. F. Carboni, T. Conner, J. Coombs, L. Daddiego, J. M. D'Ambrosio, G. Diretto, S. B. Divito, D. Douches, M. Filipiak, G. Gianese, R. Hutten, E. Jacobsen, E. Kalinska, S. Kamoun, D. Kells, H. Kossowska, L. Lopez, M. Magallanes-Lundback, T. Miranda, P. S. Naik, A. N. Pantelieva, D. Pattanayak, E. O. Patutina, M. Portantier, S. Rawat, R. Simon, B. P. Singh, B. Singh, W. Stiekema, M. V. Sukhacheva and C. Town in providing plant material, generating data, annotation, analyses, and discussions. We are indebted to additional faculty and staff of the BGI-Shenzhen, J Craig Venter Institute, and MSU Research Technology Support Facility who contributed to this project. Background and preliminary data were provided by the Centre for BioSystems Genomics (CBSG), EU-project (APOPHYS EU-QLRT-2001-01849) and US Department of Agriculture National Institute of Food and Agriculture SolCAMP project (2008-55300-04757 and 2009-85606-05673). We acknowledge the funding made available by the “863” National High Technology Development Program in China (2006AA100107), “973” National Key Basic Research Program in China (2006CB101904, 2007CB815703, 2007CB815705, 2009CB119000), Board of Wageningen University and Research Centre, CAPES - Brazilian Ministry of Education, Chinese Academy of Agricultural Sciences (seed grant to S.H.), Chinese Ministry of Agriculture (The “948” Program), Chinese Ministry of Finance (1251610601001), Chinese Ministry of Science and Technology (2007DFB30080), China Postdoctoral Science Foundation (20070420446 to Z.Z.), CONICET (Argentina), DAFF Research Stimulus Fund (07-567), CONICYT-Chile (PBCT-PSD-03), Danish Council for Strategic Research Programme Commission on Health, Food and Welfare (2101-07-0116), Danish Council for Strategic Research Programme Commission on Strategic Growth Technologies (Grant 2106-07-0021), FINCYT ((099-FINCYT-EQUIP-2009)/(076-FINCYT-PIN-2008), Préstamo BID no. 1663/OC-PE, FONDAP and BASAL-CMM), Fund for Economic Structural Support (FES), HarvestPlus Challenge Program, Indian Council of Agricultural Research, INIA-Ministry of Agriculture of Chile, Instituto Nacional de Innovación Agraria-Ministry of Agriculture of Peru, Instituto Nacional de Tecnología Agropecuaria (INTA), Italian Ministry of Research (Special Fund for Basic Research), International Potato Center (CIP-CGIAR core funds), LBMG of Center for Genome Regulation and Center for Mathematical Modeling, Universidad de Chile (UMI 2807 CNRS), Ministry of Education and Science of Russia (contract 02.552.11.7073), National Nature Science Foundation of China (30671319, 30725008, 30890032, 30971995), Natural Science Foundation of Shandong Province in China (Y2006D21), Netherlands Technology Foundation (STW), Netherlands Genomics Initiative (NGI), Netherlands Ministries of Economic Affairs (EZ) and Agriculture (LNV), New Zealand Institute for Crop & Food Research Ltd



Strategic Science Initiative, Perez Guerrero Fund, Peruvian Ministry of Agriculture-Technical Secretariat of coordination with the CGIAR, Peruvian National Council of Science and Technology (CONCYTEC), Polish Ministry of Science and Higher Education (47/PGS/2006/01), Programa Cooperativo para el Desarrollo Tecnológico Agroalimentario y Agroindustrial del Cono Sur (PROCIUSUR), Project Programa Bicentenario de Ciencia y Tecnología - Conicyt, PBCT - Conicyt PSD-03, Russian Foundation for Basic Research (09-04-12275), Secretaría de Ciencia y Tecnología (SECYT) actual Ministerio de Ciencia y Tecnología (MINCYT), Argentina, Shenzhen Municipal Government of China (CXB200903110066A, ZYC200903240077A, ZYC200903240076A), Solexa project (272-07-0196), Special Multilateral Fund of the Inter-American Council for Integral Development (FEMCIDI), Teagasc, Teagasc Walsh Fellowship Scheme, The New Zealand Institute for Plant & Food Research Ltd Capability Fund, UK Potato Genome Sequencing grant (Scottish Government Rural and Environment Research and Analysis Directorate (RERAD), Department for Environment, Food and Rural Affairs (DEFRA), Agriculture and Horticulture Development Board - Potato Council), UK Biotechnology and Biological Sciences Research Council (Grant BB/F012640), US National Science Foundation Plant Genome Research Program (DBI-0604907 / DBI-0834044), Virginia Agricultural Experiment Station USDA Hatch Funds (135853), and Wellcome Trust Strategic award (WT 083481).

**Author Contributions** A.D.G., A.G., A.N.M., A.V.B., A.V.M., B.B.K., B.K., B.R.W., B.S., B.T.L.H., B.V., B.X., B.Z., C.L., C.R.B., C.W.B.B., D.F.M., D. Martinez, D. Milbourne, D.M.A.M., D.M.B., D.D., D.M., E.D., F.G., G.A.M., G.A.T., G.D.I.C., G.G., G.J. Bishop, G.J. Bryan, G.L. G.O., G.P., G.Z., H.K., H.L., H.V.E., I.N., J.d.B., J.G., J.H., J.J., J.M.E.J., J.W., J.X., K.L.N., K.O'B., L.D., L.E.B., M.B., M.D., M.d.R.H., M.F., M. Geoffroy, M.Ghislain, M.I., M.P., M.S., M.T., N.M., N.V.R., O.P., P.F., P.N., P.S., Q.H., R.C.H.J.v.H., R.E.V., R.G., R.G.F.V., R. Lozano, R. Li, S.C., S.E.F., S.H., S.J.T., S.K.C., S.K.S., S.L., S.P., S.Y., T.B., T.V.K., V.U.P., X. Xiong, X. Xu, Y.D., Y.H., Y.L., Y.Y., Y.Z. and Z.Z. were involved in experimental design, data generation and/or data analysis. A.N.M., B.K., C.R.B., C.W.B.B., D.D., D. Milbourne, D.M.A.M., D.M.B., E.D., G.G., G.J. Bishop, G.J. Bryan, G.O., H.L., I.N., J.d.B., J.J., J.M.E.J., K.L.N., M.B., M.F., M.D., M.S., O.P., R.C.H.J.v.H., R.E.V., R.G.F.V., R. Lozano, R.W., S.E.F., S.H., S.J.T., S.K.S., T.B. and X. Xu wrote the manuscript. B.S., C.R.B., C.W.B.B., D.F.M., D. Milbourne, D.M.A.M., D.Q., G.G., G.J. Bishop, G.J. Bryan, G.O., G.P., J.M.E.J., J.W., K.G.S., R.G.F.V., R. Li, R.W., S.E.F., S.H., S.K.C., S.Y., W.Z. and Y.D. supervised data generation/analysis and managed the project. C.R.B., C.W.B.B., G.J. Bryan, G.O., J.M.E.J. and S.H. are members of The Potato Genome Sequencing Consortium Steering Committee.

**Author Information** BAC and fosmid end sequences have been deposited in the GSS division of GenBank (BAC: GS025503–GS026177, GS262924–GS365942, GS504213–GS557003; fosmid: FI900795–FI901529, FI907952–FI927051, GS557234–GS594339, GS635316–GS765761). DM Illumina GA2 WGS and Roche 454 sequences have been deposited in the NCBI Sequence Read Archive (SRA029323) and EBI Short Read Archive (ERP000411) respectively. RH NGS sequences have been deposited in the EBI Short Read Archive (ERP000627). DM and RH RNA-Seq reads have been deposited in the NCBI Sequence Read Archive (SRA030516; study SRP005965) and the European Nucleotide Database ArrayExpress Database (E-MTAB-552; study ERP000527), respectively. The DM Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AEWC01000000. The version described in this paper is the First Version, AEWC01000000. Genome sequence and annotation can be obtained and viewed at <http://potatogenome.net>. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at [www.nature.com/nature](http://www.nature.com/nature). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.H. ([huangsanwen@caas.net.cn](mailto:huangsanwen@caas.net.cn)), C.R.B. ([buell@msu.edu](mailto:buell@msu.edu)) or R.G.F.V. ([Richard.Visser@wur.nl](mailto:Richard.Visser@wur.nl)).

**The Potato Genome Consortium** (Participants are listed alphabetically by institution.)

**BGI-Shenzhen** Xun Xu<sup>1</sup>, Shengkai Pan<sup>1</sup>, Shifeng Cheng<sup>1</sup>, Bo Zhang<sup>1</sup>, Desheng Mu<sup>1</sup>, Peixiang Ni<sup>1</sup>, Gengyun Zhang<sup>1</sup>, Shuang Yang (Principal Investigator)<sup>1</sup>, Ruiqiang Li (Principal Investigator)<sup>1</sup>, Jun Wang (Principal Investigator)<sup>1</sup>; **Cayetano Heredia University** Gisella Orjeda (Principal Investigator)<sup>2</sup>, Frank Guzman<sup>2</sup>, Michael Torres<sup>2</sup>, Roberto Lozano<sup>2</sup>, Olga Ponce<sup>2</sup>, Diana Martinez<sup>2</sup>, Germán De la Cruz<sup>2</sup>; **Central Potato Research Institute** S. K. Chakrabarti (Principal Investigator)<sup>3</sup>, Virupaksh U. Patil<sup>3</sup>; **Centre Bioengineering RAS** Konstantin G. Skryabin (Principal Investigator)<sup>4</sup>, Boris B. Kuznetsov<sup>4</sup>, Nikolai V. Ravin<sup>4</sup>, Tatjana V. Kolganova<sup>4</sup>, Alexey V. Beletsky<sup>4</sup>, Andrei V. Mardanov<sup>4</sup>; **CGR-CMM, Universidad de Chile** Alex Di Genova<sup>5</sup>; **College of Life Sciences, University of Dundee** Daniel M. Bolser<sup>6</sup>, David M. A. Martin (Principal Investigator)<sup>6</sup>; **High Technology Research Center, Shandong Academy of Agricultural Sciences** Guangcun Li<sup>7</sup>, Yu Yang<sup>7</sup>; **Huazhong Agricultural University** Hanhui Kuang<sup>8</sup>, Qun Hu<sup>8</sup>; **Hunan Agricultural University** Xingyao Xiong<sup>9</sup>; **Imperial College London**

Gerard J. Bishop<sup>10</sup>; **Instituto de Investigaciones Agropecuarias** Boris Sagredo (Principal Investigator)<sup>11</sup>, Nilo Mejía<sup>11</sup>; **Institute of Biochemistry & Biophysics** Włodzimierz Zagorski (Principal Investigator)<sup>12</sup>, Robert Gromadka<sup>12</sup>, Jan Gawor<sup>12</sup>, Paweł Szczesny<sup>12</sup>; **Institute of Vegetables & Flowers, Chinese Academy of Agricultural Sciences** Sanwen Huang (Principal Investigator)<sup>13</sup>, Zhonghua Zhang<sup>13</sup>, Chunbo Liang<sup>13</sup>, Jun He<sup>13</sup>, Ying Li<sup>13</sup>, Ying He<sup>13</sup>, Jianfei Xu<sup>13</sup>, Youjun Zhang<sup>13</sup>, Binyan Xie<sup>13</sup>, Yongchen Du<sup>13</sup>, Dongyu Qu (Principal Investigator)<sup>13</sup>; **International Potato Center** Merideth Bonierbale<sup>14</sup>, Marc Ghislain<sup>14</sup>, Maria del Rosario Herrera<sup>14</sup>; **Italian National Agency for New Technologies, Energy & Sustainable Development** Giovanni Giuliano (Principal Investigator)<sup>15</sup>, Marco Pietrella<sup>15</sup>, Gaetano Perrotta<sup>15</sup>, Paolo Facella<sup>15</sup>; **J Craig Venter Institute** Kimberly O'Brien<sup>16</sup>; **Laboratorio de Agrobiotecnología, Instituto Nacional de Tecnología Agropecuaria** Sergio E. Feingold (Principal Investigator)<sup>17</sup>, Leandro E. Barreiro<sup>17</sup>, Gabriela A. Massa<sup>17</sup>; **Laboratorio de Biología de Sistemas, Universidad Nacional de La Plata** Luis Diambra<sup>18</sup>; **Michigan State University** Brett R. Whitty<sup>19</sup>, Brieanne Vaillancourt<sup>19</sup>, Haining Lin<sup>19</sup>, Alicia N. Massa<sup>19</sup>, Michael Geoffroy<sup>19</sup>, Steven Lundback<sup>19</sup>, Dean DellaPenna<sup>19</sup>, C. Robin Buell (Principal Investigator)<sup>19</sup>; **Scottish Crop Research Institute** Sanjeev Kumar Sharma<sup>20†</sup>, David F. Marshall<sup>20†</sup>, Robbie Waugh<sup>20†</sup>, Glenn J. Bryan (Principal Investigator)<sup>20†</sup>; **Teagasc Crops Research Centre** Marialaura Destefanis<sup>21</sup>, Istvan Nagy<sup>21</sup>, Dan Milbourne (Principal Investigator)<sup>21</sup>; **The New Zealand Institute for Plant & Food Research Ltd** Susan J. Thomson<sup>22</sup>, Mark Fiers<sup>22</sup>, Jeanne M. E. Jacobs (Principal Investigator)<sup>22</sup>; **University of Aalborg** Kåre L. Nielsen (Principal Investigator)<sup>23</sup>, Mads Sønderkær<sup>23</sup>; **University of Wisconsin** Marina Iovene<sup>24</sup>, Giovana A. Torres<sup>24</sup>, Jiming Jiang (Principal Investigator)<sup>24</sup>; **Virginia Polytechnic Institute & State University** Richard E. Veilleux<sup>25</sup>; **Wageningen University & Research Center** Christian W. B. Bachem (Principal Investigator)<sup>26</sup>, Jan de Boer<sup>26</sup>, Theo Borm<sup>26</sup>, Bjorn Kloosterman<sup>26</sup>, Herman van Eck<sup>26</sup>, Erwin Datema<sup>27</sup>, Bas te Lintel Hekkert<sup>27</sup>, Aska Goverse<sup>28,29</sup>, Roeland C. H. J. van Ham<sup>27,28</sup> & Richard G. F. Visser<sup>26,28</sup>

<sup>1</sup>BGI-Shenzhen, Chinese Ministry of Agricultural, Key Lab of Genomics, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China. <sup>2</sup>Cayetano Heredia University, Genomics Research Unit, Av Honorio Delgado 430, Lima 31, Peru and San Cristobal of Huamanga University, Biotechnology and Plant Genetics Laboratory, Ayacucho, Peru. <sup>3</sup>Central Potato Research Institute, Shimla 171001, Himachal Pradesh, India. <sup>4</sup>Centre Bioengineering RAS, Prospekt 60-letya Oktyabrya, 7-1, Moscow 117312, Russia. <sup>5</sup>Center for Genome Regulation and Center for Mathematical Modeling, Universidad de Chile (UMI 2807 CNRS), Chile. <sup>6</sup>College of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, UK. <sup>7</sup>High Technology Research Center, Shandong Academy of Agricultural Sciences, 11 Sangyuan Road, Jinan 250100, P. R. China. <sup>8</sup>Huazhong Agriculture University, Ministry of Education, College of Horticulture and Forestry, Department of Vegetable Crops, Key Laboratory of Horticulture Biology, Wuhan 430070, P. R. China. <sup>9</sup>Hunan Agricultural University, College of Horticulture and Landscape, Changsha, Hunan 410128, China. <sup>10</sup>Imperial College London, Division of Biology, South Kensington Campus, London SW7 1AZ, UK. <sup>11</sup>Instituto de Investigaciones Agropecuarias, Avda. Salamanca s/n, Km 105 ruta 5 sur, sector Los Choapiños. Rengo, Región del Libertador Bernardo O'Higgins, Código Postal 2940000, Chile. <sup>12</sup>Institute of Biochemistry and Biophysics, DNA Sequencing and Oligonucleotides Synthesis Laboratory, PAS ul. Pawinskiego 5a, 02-106 Warsaw, Poland. <sup>13</sup>Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Key Laboratory of Horticultural Crops Genetic Improvement of Ministry of Agriculture, Sino-Dutch Joint Lab of Horticultural Genomics Technology, Beijing 100081, China. <sup>14</sup>International Potato Center, P.O. Box 1558, Lima 12, Peru. <sup>15</sup>Italian National Agency for New Technologies, Energy and Sustainable Development (ENEA), Casaccia Research Center, Via Anguillarese 301, 00123 Roma, Italy and Trisaia Research Center, S.S. 106 Ionica - Km 419.50 75026 Rotondella (Matera), Italy. <sup>16</sup>J Craig Venter Institute, 9712 Medical Center Dr, Rockville, Maryland 20850, USA. <sup>17</sup>Laboratorio de Agrobiotecnología, Estación Experimental Agropecuaria Balcarce, Instituto Nacional de Tecnología Agropecuaria (INTA) cc276 (7620) Balcarce, Argentina. <sup>18</sup>Laboratorio de Biología de Sistemas, CREG, Universidad Nacional de La Plata, 1888, Argentina. <sup>19</sup>Michigan State University, East Lansing, Michigan 48824, USA. <sup>20</sup>Scottish Crop Research Institute, Genetics Programme, Invergowrie, Dundee DD2 5DA, UK. <sup>21</sup>Teagasc Crops Research Centre, Oak Park, Carlow, Ireland. <sup>22</sup>The New Zealand Institute for Plant & Food Research Ltd., Private Bag 4704, Christchurch 8140, New Zealand. <sup>23</sup>University of Aalborg (AAU), Department of Biotechnology, Chemistry and Environmental Engineering, Sohngaardsholmsvej 49, 9000 Aalborg, Denmark. <sup>24</sup>University of Wisconsin-Madison, Department of Horticulture, 1575 Linden Drive, Madison, Wisconsin 53706, USA. <sup>25</sup>Virginia Polytechnic Institute and State University, Department of Horticulture, 544 Latham Hall, Blacksburg, Virginia 24061, USA. <sup>26</sup>Wageningen University and Research Center, Dept. of Plant Sciences, Laboratory of Plant Breeding, Droevendaalsesteeg 1, 6708PB Wageningen, Netherlands. <sup>27</sup>Wageningen University and Research Center, Applied Bioinformatics, Plant Research International, Droevendaalsesteeg 1, 6708PB Wageningen, Netherlands. <sup>28</sup>Centre for BioSystems Genomics, Droevendaalsesteeg 1, 6708PB Wageningen, Netherlands. <sup>29</sup>Wageningen University and Research Center, Dept. of Plant Sciences, Laboratory of Nematology, Droevendaalsesteeg 1, 6708PB Wageningen, Netherlands. †Present address: The James Hutton Institute, Invergowrie, Dundee, DD2 5DA, UK (S.K.S., D.F.M., R.W., G.J. Bryan).

## METHODS

**DM whole-genome shotgun sequencing and assembly.** Libraries were constructed from DM genomic DNA and sequenced on the Sanger, Illumina Genome Analyser 2 (GA2) and Roche 454 platforms using standard protocols (see Supplementary Text). A BAC library and three fosmid libraries were end sequenced using the Sanger platform. For the Illumina GA2 platform, we generated 70.6 Gb of 37–73 bp paired-end reads from 16 libraries with insert lengths of 200–811 bp (Supplementary Tables 21 and 22). We also generated 18.7 Gb of Illumina mate-pair libraries (2, 5 and 10 kb insert size). In total, 7.2 Gb of 454 single-end data were generated and applied for gap filling to improve the assembly, of which 4.7 Gb (12,594,513 reads) were incorporated into the final assembly. For the 8 and 20 kb 454 paired-end reads, representing 0.7 and 1.0 Gb of raw data respectively, 90.7 Mb (511,254 reads) and 211 Mb (1,525,992 reads), respectively, were incorporated into the final assembly.

We generated a high-quality potato genome using the short read assembly software SOAPdenovo<sup>4</sup> (Version 1014). We first assembled 69.4 Gb of GA2 paired-end short reads into contigs, which are sequence assemblies without gaps composed of overlapping reads. To increase the assembly accuracy, only 78.3% of the reads with high quality were considered. Then contigs were further linked into scaffolds by paired-end relationships (~300 to ~550 bp insert size), mate-pair reads (2 to approximately 10 kb), fosmid ends (~40 kb, 90,407 pairs of end sequences) and BAC ends (~100 kb, 71,375 pairs of end sequences). We then filled gaps with the entire short-read data generated using Illumina GA2 reads. The primary contig  $N_{50}$  size (the contig length such that using equal or longer contigs produces half of the bases of the assembled genome) was 697 bp and increased to 1,318 kb after gap-filling (Supplementary Tables 23 and 24). When only the paired-end relationships were used in the assembly process, the  $N_{50}$  scaffold size was 22.4 kb. Adding mate-pair reads with 2, 5 and 10 kb insert sizes, the  $N_{50}$  scaffold size increased to 67, 173 and 389 kb, respectively. When integrated with additional libraries of larger insert size, such as fosmid and BAC end sequences, the  $N_{50}$  reached 1,318 kb. The final assembly size was 727 Mb, 93.87% of which is non-gapped sequence. We further filled the gaps with 6.74 fold coverage of 454 data, which increased the  $N_{50}$  contig size to 31,429 bp with 15.4% of the gaps filled.

The single-base accuracy of the assembly was estimated by the depth and proportion of discordant reads. For the DM v3.0 assembly, 95.45% of 880 million usable reads could be mapped back to the assembled genome by SOAP 2.20 (ref. 34) using optimal parameters. The read depth was calculated for each genomic location and peak depth for whole genome and the CDS regions are 100 and 105, respectively. Approximately 96% of the assembled sequences had more than 20-fold coverage (Supplementary Fig. 1). The overall GC content of the potato genome is about 34.8% with a positive correlation between GC content and sequencing depth (data not shown). The DM potato should have few heterozygous sites and 93.04% of the sites can be supported by at least 90% reads, suggesting high base quality and accuracy.

**RH genome sequencing.** Whole-genome sequencing of genotype RH was performed on the Illumina GA2 platform using a variety of fragment sizes and reads lengths resulting in a total of 144 Gb of raw data (Supplementary Table 25). These data were filtered using a custom C program and assembled using SOAPdenovo 1.03 (ref. 4). Additionally, four 20-kb mate-pair libraries were sequenced on a Roche 454 Titanium sequencer, amounting to 581 Mb of raw data (Supplementary Table 26). The resulting sequences were filtered for duplicates using custom Python scripts.

The RH BACs were sequenced using a combination of Sanger and 454 sequencing at various levels of coverage (Supplementary Tables 9–11). Consensus base calling errors in the BAC sequences were corrected using custom Python and C scripts using a similar approach to that described previously<sup>35</sup> (Supplementary Text). Sequence overlaps between BACs within the same physical tiling path were identified using megablast from BLAST 2.2.21 (ref. 36) and merged with megamerger from the EMBOSS 6.1.0 package<sup>37</sup>. Using the same pipeline, several kilobase-sized gaps were closed through alignment of a preliminary RH whole-genome assembly. The resulting non-redundant contigs were scaffolded by mapping the RH whole-genome Illumina and 454 mated sequences against these contigs using SOAPalign 2.20 (ref. 34) and subsequently processing these mapping results with a custom Python script. The scaffolds were then ordered into superscaffolds based on the BAC order in the tiling paths of the FPC map. This procedure removed 25 Mb of redundant sequence, reduced the number of sequence fragments from 17,228 to 3,768, and increased the  $N_{50}$  sequence length from 24 to 144 kb (Supplementary Tables 9 and 10).

**Construction of the DM genetic map and anchoring of the genome.** To anchor and fully orientate physical contigs along the chromosome, a genetic map was developed *de novo* using sequence-tagged-site (STS) markers comprising simple sequence repeats (SSR), SNPs, and diversity array technology (DaT). SSR and

SNP markers were designed directly from assembled sequence scaffolds, whereas polymorphic DaT marker sequences were searched against the scaffolds for high-quality unique matches. A total of 4,836 STS markers including 2,174 DaTs, 2,304 SNPs and 358 SSRs were analysed on 180 progeny clones from a backcross population ((DM × DI) × DI) developed at CIP between DM and DI (CIP no. 703825), a heterozygous diploid *S. tuberosum* group Stenotomum (formerly *S. stenotomum* ssp. *goniocalyx*) landrace clone. The data from 2,603 polymorphic STS markers comprising 1,881 DaTs, 393 SNPs and 329 SSR alleles were analysed using JoinMap 4 (ref. 38) and yielded the expected 12 potato linkage groups. Supplementary Fig. 3 represents the mapping and anchoring of the potato genome, using chromosome 7 as an example.

Anchoring the DM genome was accomplished using direct and indirect approaches. The direct approach employed the ((DM × DI) × DI) linkage map whereby 2,037 of the 2,603 STS markers comprised of 1,402 DaTs, 376 SNPs and 259 SSRs could be uniquely anchored on the DM superscaffolds. This approach anchored ~52% (394 Mb) of the assembly arranged into 334 superscaffolds (Supplementary Table 27 and Supplementary Fig. 3).

RH is the male parent of the mapping population of the ultra-high-density (UHD) linkage map<sup>28</sup> used for construction and genetic anchoring of the physical map using the RHPOTKEY BAC library<sup>39</sup>. The indirect mapping approach exploited *in silico* anchoring using the RH genetic and physical map<sup>28,40</sup>, as well as tomato genetic map data from SGN (<http://solgenomics.net/>). Amplified fragment length polymorphism markers from the RH genetic map were linked to DM sequence scaffolds via BLAST alignment<sup>36</sup> of whole-genome-profiling sequence tags<sup>41</sup> obtained from anchored seed BACs in the RH physical map, or by direct alignment of fully sequenced RH seed BACs to the DM sequence. The combined marker alignments were processed into robust anchor points. The tomato sequence markers from the genetic maps were aligned to the DM assembly using SSAHA2 (ref. 42). Positions of ambiguously anchored superscaffolds were manually checked and corrected. This approach anchored an additional ~32% of the assembly (229 Mb). In 294 cases, the two independent approaches provided direct support for each other, anchoring the same scaffold to the same position on the two maps.

Overall, the two strategies anchored 649 superscaffolds to approximate positions on the genetic map of potato covering a length of 623 Mb. The 623 Mb (~86%) anchored genome includes ~90% of the 39,031 predicted genes. Of the unanchored superscaffolds, 84 were found in the N90 (622 scaffolds greater than 0.25 Mb), constituting 17 Mb of the overall assembly or 2% of the assembled genome. The longest anchored superscaffold is 7 Mb (from chromosome 1) and the longest unanchored superscaffold is 2.5 Mb.

**Identification of repetitive sequences.** Transposable elements (TEs) in the potato genome assembly were identified at the DNA and protein level. RepeatMasker<sup>29</sup> was applied using Repbase<sup>43</sup> for TE identification at the DNA level. At the protein level, RepeatProteinMask<sup>29,44</sup> was used in a WuBlastX<sup>36</sup> search against the TE protein database to further identify TEs. Overlapping TEs belonging to the same repeat class were collated, and sequences were removed if they overlapped >80% and belonged to different repeat classes.

**Gene prediction.** To predict genes, we performed *ab initio* predictions on the repeat-masked genome and then integrated the results with spliced alignments of proteins and transcripts to genome sequences using GLEAN<sup>30</sup>. The potato genome was masked by identified repeat sequences longer than 500 bp, except for miniature inverted repeat transposable elements which are usually found near genes or inside introns<sup>45</sup>. The software Augustus<sup>46</sup> and Genscan<sup>47</sup> was used for *ab initio* predictions with parameters trained for *A. thaliana*. For similarity-based gene prediction, we aligned the protein sequences of four sequenced plants (*A. thaliana*, *Carica papaya*, *V. vinifera* and *Oryza sativa*) onto the potato genome using TBLASTN with an *E*-value cut-off of  $1 \times 10^{-5}$ , and then similar genome sequences were aligned against the matching proteins using Genewise<sup>48</sup> for accurately spliced alignments. In EST-based predictions, EST sequences of 11 *Solanum* species were aligned against the potato genome using BLAT (identity  $\geq 0.95$ , coverage  $\geq 0.90$ ) to generate spliced alignments. All these resources and prediction approaches were combined by GLEAN<sup>30</sup> to build the consensus gene set. To finalize the gene set, we aligned the RNA-Seq from 32 libraries, of which eight were sequenced with both single- and paired-end reads, to the genome using Tophat<sup>31</sup> and the alignments were then used as input for Cufflinks<sup>32</sup> using the default parameters. Gene, transcript and peptide sets were filtered to remove small genes, genes modelled across sequencing gaps, TE-encoding genes, and other incorrect annotations. The final gene set contains 39,031 genes with 56,218 protein-coding transcripts, of which 52,925 nonidentical proteins were retained for analysis.

**Transcriptome sequencing.** RNA was isolated from many tissues of DM and RH that represent developmental, abiotic stress and biotic stress conditions (Supplementary Table 4 and Supplementary Text). cDNA libraries were constructed (Illumina) and sequenced on an Illumina GA2 in the single- and/or paired-end



mode. To represent the expression of each gene, we selected a representative transcript from each gene model by selecting the longest CDS from each gene. The aligned read data were generated by Tophat<sup>31</sup> and the selected transcripts used as input into Cufflinks<sup>32</sup>, a short-read transcript assembler that calculates the fragments per kb per million mapped reads (FPKM) as expression values for each transcript. Cufflinks was run with default settings, with a maximum intron length of 15,000. FPKM values were reported and tabulated for each transcript (Supplementary Table 19).

**Comparative genome analyses.** Paralogous and orthologous clusters were identified using OrthoMCL<sup>49</sup> using the predicted proteomes of 11 plant species (Supplementary Table 28). After removing 1,602 TE-related genes that were not filtered in earlier annotation steps, asterid-specific and potato-lineage-specific genes were identified using the initial OrthoMCL clustering followed by BLAST searches (*E*-value cut-off of  $1 \times 10^{-5}$ ) against assemblies of ESTs available from the PlantGDB project (<http://plantgdb.org>; 153 nonasterid species and 57 asterid species; Supplementary Fig. 5 and Supplementary Table 29). Analysis of protein domains was performed using the Pfam hmm models identified by InterProScan searches against InterPro (<http://www.ebi.ac.uk/interpro>). We compared the Pfam domains of the asterid-specific and potato-lineage-specific sets with those that are shared with at least one other nonasterid genome or transcriptome. A Fisher's exact test was used to detect significant differences in Pfam representation between protein sets.

After removing the self and multiple matches, the syntenic blocks ( $\geq 5$  genes per block) were identified using MCScan<sup>9</sup> and i-adhore 3.0 (ref. 50) based on the aligned protein gene pairs (Supplementary Table 8). For the self-aligned results, each aligned block represents the paralogous segments pair that arose from the genome duplication whereas, for the inter-species alignment results, each aligned block represents the orthologous pair derived from the shared ancestor. We calculated the 4DTV (fourfold degenerate synonymous sites of the third codons) for each gene pair from the aligned block and give a distribution for the 4DTV value to estimate the speciation or WGD event that occurred in evolutionary history.

**Identification of disease resistance genes.** Predicted open reading frames (ORFs) from the annotation of *S. tuberosum* group Phureja assembly V3 were screened using HMMER V.3 (<http://hmmerr.janelia.org/software>) against the raw hidden Markov model (HMM) corresponding to the Pfam NBS (NB-ARC) family (PF00931). The HMM was downloaded from the Pfam home page (<http://pfam.sanger.ac.uk/>). The analysis using the raw HMM of the NBS domain resulted in 351 candidates. From these, a high quality protein set ( $< 1 \times 10^{-60}$ ) was aligned and used to construct a potato-specific NBS HMM using the module 'hmmbuild'. Using this new potato-specific model, we identified 500 NBS-candidate proteins that were individually analysed. To detect TIR and LRR domains, Pfam HMM searches were used. The raw TIR HMM (PF01582) and LRR 1 HMM (PF00560) were downloaded and compared against the two sets of NBS-encoding amino acid sequences using HMMER V3. Both TIR and LRR domains were validated using NCBI conserved domains and multiple expectation maximization for motif elicitation (MEME)<sup>51</sup>. In the case of LRRs, MEME was also useful to detect the number of repeats of this particular domain in the protein. As previously reported<sup>52</sup>, Pfam analysis could not identify the CC motif in the N-terminal region. CC domains were thus analysed using the MARCOIL<sup>53</sup> program with a threshold probability of 90 (ref. 52) and double-checked using paircoil2 (ref. 54) with a *P*-score cut-off of 0.025 (ref. 55). Selected genes ( $\pm 1.5$  kb) were searched using BLASTX against a reference *R*-gene set<sup>56</sup> to find a well-characterized homologue. The reference set was used to select and annotate as pseudogenes those peptides that had large deletions, insertions, frameshift mutations, or premature stop codons. DNA and protein comparisons were used.

**Haplotype diversity analysis.** RH reads generated by the Illumina GA2 were mapped onto the DM genome assembly using SOAP2.20 (ref. 34) allowing at most four mismatches and SNPs were called using SOAPsnp. Q20 was used to filter the SNPs owing to sequencing errors. To exclude SNP calling errors caused by incorrect alignments, we excluded adjacent SNPs separated by  $< 5$  bp. SOAPindel was used to detect the indels between DM and RH. Only indels supported by more than three uniquely mapped reads were retained. Owing to the heterozygosity of RH, the SNPs and indels were classified into heterozygous and homozygous SNPs or indels.

On the basis of the annotated genes in the DM genome assembly, we extracted the SNPs located at coding regions and stop codons. If a homozygous SNP in RH within a coding region induced a premature stop codon, we defined the gene harbouring this SNP as a homozygous premature stop gene in RH. If the SNP inducing a premature stop codon was heterozygous, the gene harbouring this

SNP was considered a heterozygous premature stop codon gene in RH. In addition, both categories can be further divided into premature stop codons shared with DM or not shared with DM. As a result, the numbers of premature stop codons are 606 homozygous PS genes in RH, 1,760 heterozygous PS genes in RH but not shared with DM, 288 PS in DM only, and 652 heterozygous premature stop codons in RH and shared by DM.

To identify genes with frameshift mutations in RH, we identified all the genes containing indels of which the length could not be divided by 3. We found 80 genes with frameshift mutations, of which 31 were heterozygous and 49 were homozygous.

To identify DM-specific genes, we mapped all the RH Illumina GA2 reads to the DM genome assembly. If the gene was not mapped to any RH read, it was considered a DM-specific gene. We identified 35 DM-specific genes, 11 of which are supported by similarity to entries in the KEGG database<sup>57</sup>. To identify RH-specific genes, we assembled the RH Illumina GA2 reads that did not map to the DM genome into RH-specific scaffolds. Then, these scaffolds were annotated using the same strategy as for DM. To exclude contamination, we aligned the CDS sequences against the protein set of bacteria with the *E*-value cut-off of  $1 \times 10^{-5}$  using Blastx. CDS sequences with  $> 90\%$  identity and  $> 90\%$  coverage were considered contaminants and were excluded. In addition, all DM RNA-seq reads were mapped onto the CDS sequences, and CDS sequences with homologous reads were excluded because these genes may be due to incorrect assembly. In total, we predicted 246 RH specific genes, 34 of which are supported by Gene Ontology annotation<sup>17</sup>.

34. Li, R. *et al.* SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**, 1966–1967 (2009).
35. Chaisson, M., Pevzner, P. & Tang, H. Fragment assembly with short reads. *Bioinformatics* **20**, 2067–2074 (2004).
36. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
37. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
38. Van Ooijen, J. W. in *JoinMap 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations* (ed. Kyazma, B. V.) (Wageningen, 2006).
39. Borm, T. J. *Construction and Use of a Physical Map of Potato*. PhD thesis, Wageningen Univ. (2008).
40. Visser, R. G. F. *et al.* Sequencing the potato genome: outline and first results to come from the elucidation of the sequence of the world's third most important crop. *Am. J. Potato Res.* **86**, 417–429 (2009).
41. Van der Vossen, E. *et al.* in *Whole Genome Profiling of the Diploid Potato Clone RH89-039-16* (Plant & Animal Genomes XVIII Conference, 2010).
42. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: a fast search method for large DNA databases. *Genome Res.* **11**, 1725–1729 (2001).
43. Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
44. Jiang, Z., Hubley, R., Smit, A. & Eichler, E. E. DupMasker: a tool for annotating primate segmental duplications. *Genome Res.* **18**, 1362–1368 (2008).
45. Kuang, H. *et al.* Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: new functional implications for MITes. *Genome Res.* **19**, 42–56 (2009).
46. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**, W309–W312 (2004).
47. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
48. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004).
49. Chen, F., Mackey, A. J., Vermunt, J. K. & Roos, D. S. Assessing performance of orthology detection strategies applied to eukaryotic genomes. *PLoS ONE* **2**, e383 (2007).
50. Simillion, C., Janssens, K., Sterck, L. & Van de Peer, Y. i-ADHoRe 2.0: an improved tool to detect degenerated genomic homology using genomic profiles. *Bioinformatics* **24**, 127–128 (2008).
51. Bailey, T. L. & Elkan, C. The value of prior knowledge in discovering motifs with MEME. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **3**, 21–29 (1995).
52. Mun, J. H., Yu, H. J., Park, S. & Park, B. S. Genome-wide identification of NBS-encoding resistance genes in *Brassica rapa*. *Mol. Genet. Genomics* **282**, 617–631 (2009).
53. Delorenzi, M. & Speed, T. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **18**, 617–625 (2002).
54. McDonnell, A. V., Jiang, T., Keating, A. E. & Berger, B. Paircoil2: improved predictions of coiled coils from sequence. *Bioinformatics* **22**, 356–358 (2006).
55. Porter, B. W. *et al.* Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. *Mol. Genet. Genomics* **281**, 609–626 (2009).
56. Sanseverino, W. *et al.* PRGdb: a bioinformatics platform for plant resistance gene analysis. *Nucleic Acids Res.* **38**, D814–D821 (2010).
57. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, D277–D280 (2004).

# Functional regeneration of respiratory pathways after spinal cord injury

Warren J. Alilain<sup>1</sup>, Kevin P. Horn<sup>1</sup>, Hongmei Hu<sup>1</sup>, Thomas E. Dick<sup>1,2</sup> & Jerry Silver<sup>1</sup>

**Spinal cord injuries often occur at the cervical level above the phrenic motor pools, which innervate the diaphragm. The effects of impaired breathing are a leading cause of death from spinal cord injuries, underscoring the importance of developing strategies to restore respiratory activity. Here we show that, after cervical spinal cord injury, the expression of chondroitin sulphate proteoglycans (CSPGs) associated with the perineuronal net (PNN) is upregulated around the phrenic motor neurons. Digestion of these potentially inhibitory extracellular matrix molecules with chondroitinase ABC (denoted ChABC) could, by itself, promote the plasticity of tracts that were spared and restore limited activity to the paralysed diaphragm. However, when combined with a peripheral nerve autograft, ChABC treatment resulted in lengthy regeneration of serotonin-containing axons and other bulbospinal fibres and remarkable recovery of diaphragmatic function. After recovery and initial transection of the graft bridge, there was an unusual, overall increase in tonic electromyographic activity of the diaphragm, suggesting that considerable remodelling of the spinal cord circuitry occurs after regeneration. This increase was followed by complete elimination of the restored activity, proving that regeneration is crucial for the return of function. Overall, these experiments present a way to markedly restore the function of a single muscle after debilitating trauma to the central nervous system, through both promoting the plasticity of spared tracts and regenerating essential pathways.**

CSPGs are a key component of PNNs and glial scars, and they powerfully inhibit neuronal plasticity, sprouting and regeneration<sup>1–4</sup>. Degradation of these inhibitors with ChABC can restore some function through increasing the regeneration of severed axons, as well as enhancing the sprouting and/or improving the conduction of spared fibres<sup>5,6</sup>. After lateral hemisection at the second cervical level of the spinal cord (C2), there is a rapid upregulation of CSPG expression in the vicinity of the lesion but also distally, at the level of the phrenic nucleus (C3–C6) (Fig. 1a and Supplementary Fig. 1). Such an increase in PNN-associated proteoglycans far from a spinal cord injury was first demonstrated in deafferented dorsal column nuclei, and we now report a similar phenomenon around denervated motor neurons<sup>4</sup>. One recently discovered mechanism that governs the upregulation of CSPGs at lesion sites is the extravasation of a complex of fibrinogen and transforming growth factor- $\beta$  through the open blood–brain barrier, triggering the release of extracellular matrix components by reactive astrocytes<sup>7</sup>. The function of PNN-associated CSPG upregulation at sites distal to lesions (other than impeding plasticity) and the mechanism that leads to this upregulation are unknown but are probably associated with pro-inflammatory stimuli in deafferented nuclei.

## Functional effects of PNN enzymatic degradation

*In vivo* administration of ChABC (250 nl of a 20 U ml<sup>-1</sup> saline solution) at the level of the phrenic nucleus significantly degraded the CSPG family of inhibitors and resulted in accumulation of the CSPG ‘stub’ antigen, as visualized by immunohistochemical staining with an antibody specific for digested CSPGs (2B6) (Fig. 1b, c). Immunohistochemistry and histology showed that at 5 weeks after C2 hemisection and administration of ChABC, the PNN had not reappeared, allowing the potential for continued sprouting (Supplementary Fig. 2a). The PNN had still not reformed at 12 weeks (data not shown) but had mostly reappeared by 5 months after ChABC administration (Supplementary Fig. 2b).

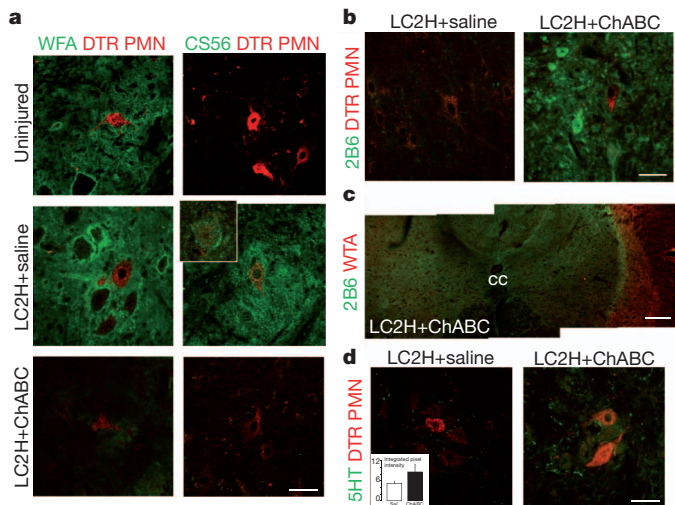
Within 1 week, and persisting over time (Supplementary Fig. 8), treatment with ChABC led to an increase in the number of serotonergic fibres surrounding the phrenic motor neurons compared with C2-hemisectioned animals that had received only saline, with pixel intensity values of stained serotonin (5-HT) almost double those of saline-treated animals (Fig. 1d). This finding is important because 5-HT has a crucial role in functional respiratory plasticity, especially under stressful conditions; it increases the efficacy of the ‘crossed phrenic pathway’: that is, of the small contingent of contralateral glutamatergic fibres from the rostral ventral respiratory group (rVRG), which innervates the phrenic nucleus and remains after C2 hemisection<sup>8–13</sup>. Indeed, electromyographic (EMG) recording showed that, when inducing the crossed phrenic phenomenon by transecting the phrenic nerve contralateral to the hemisection, which (although clinically irrelevant) markedly increases respiratory drive and activates the crossed phrenic pathway, there was an augmented return of activity in the hemidiaphragm ipsilateral to the lesion in ChABC-treated animals. The activity was at least twice that seen in phrenicotomy, non-ChABC-treated animals (Supplementary Fig. 3).

As early as 1 week after C2 hemisection, treatment with ChABC, even without a phrenicotomy, could also lead to some recovery, whereas vehicle (saline)-treated animals showed no recovery, at least at this early time point (Supplementary Fig. 4). A small minority of animals improved spontaneously, albeit minimally, over time without intervention (Fig. 2c). The recovery of EMG activity in animals treated solely with ChABC was more rapid than that of untreated animals (beginning at 1 week instead of 6–8 weeks) and occurred in a larger number of animals (73% versus 18%). However, even 12 weeks after treatment, the recovery of breathing function was meagre, and inspiratory bursts did not exceed the 10–20% peak amplitude level that could infrequently be achieved in control animals (Fig. 2c).

Because recovery was ultimately disappointing following ChABC treatment alone, we sought to improve it by implanting an autologous

<sup>1</sup>Department of Neurosciences, Case Western Reserve University School of Medicine, 2109 Adelbert Road, Cleveland, Ohio 44106, USA. <sup>2</sup>Division of Pulmonary, Critical Care, and Sleep Medicine, Department of Medicine, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106, USA.

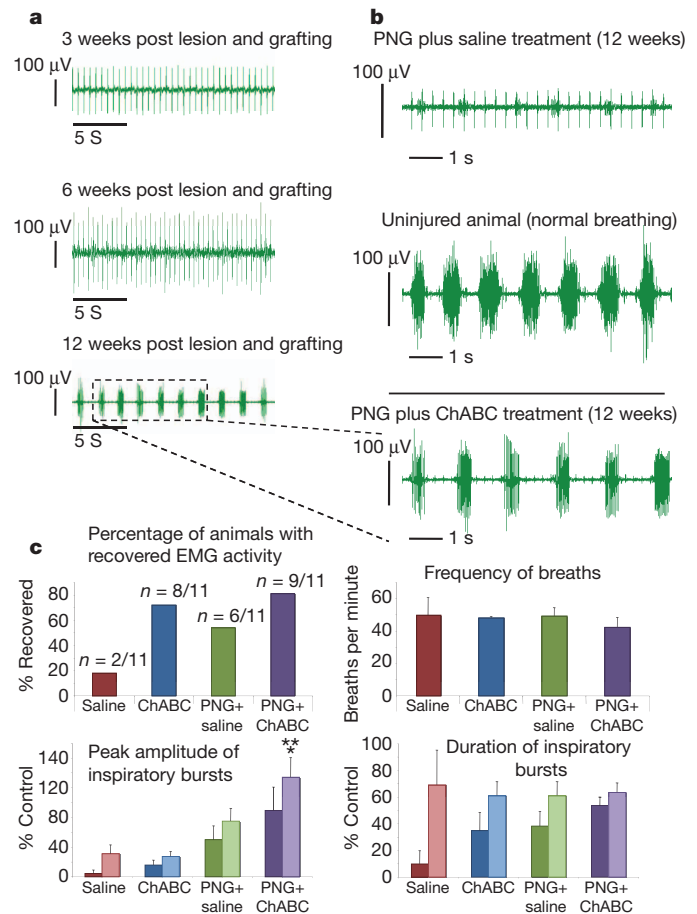




**Figure 1 | C2 hemisection results in CSPG production.** At 7 days after C2 hemisection, there is increased expression of the PNN and inhibitory CSPGs around phrenic motor neurons. **a**, In uninjured animals (top), phrenic motor neurons labelled with dextran amine Texas Red (DTR PMN; red) are ensheathed by the PNN, indicated by staining with *Wisteria floribunda* agglutinin (WFA; green, left). But these neurons stained poorly for CS56, a marker of CSPGs (green, right). After C2 hemisection and treatment with saline (centre), there is increased WFA and CS56 staining. The insert shows another example of a labelled phrenic motor neuron enveloped by CSPGs. At 7 days after ChABC treatment (bottom), WFA and CS56 staining disappears. Scale bar, 40  $\mu$ m. **b**, A marker for digested CSPGs (2B6; green) is present after ChABC treatment. Scale bar, 40  $\mu$ m. **c**, A lower-power montage shows the area of CSPG digestion by ChABC (the side of the montage left of the central canal (CC) is ipsilateral to the lesion and ChABC treatment). Scale bar, 200  $\mu$ m. **d**, Within the region of CSPG degradation, there is an increase in serotonergic fibres (5-HT; green) around phrenic motor neurons. Scale bar, 40  $\mu$ m. Pixel intensity analysis shows a doubling in the amount of 5-HT in this region in ChABC-treated animals compared with saline (sal)-treated animals ( $n = 6$ ; inset). Error bars, s.e.m.

peripheral nerve graft (PNG), whose resident Schwann cells would provide trophic support, remyelination and guided, long-distance regeneration, directly by-passing the lesion to deliver axons in the vicinity of the denervated phrenic nucleus<sup>14–18</sup> (Supplementary Fig. 5). The administration of ChABC allowed local sprouting, as well as enhanced graft entry and exit of axons<sup>19,20</sup> (Supplementary Fig. 5). At 3, 6, 9 and 12 weeks after C2 hemisection, the animals were assessed for return of diaphragmatic muscle activity by bilateral EMG recordings (Fig. 2a). At about 10 weeks (data not shown) and 12 weeks post C2 hemisection and grafting, animals that received ChABC as well as an autologous PNG showed the most recovery compared with lesioned animals that received ChABC alone, saline alone or a PNG with saline treatment (Fig. 2b, c). Furthermore, in recovered animals, the peak inspiratory amplitude of the raw EMG trace was nearly equivalent to, and often surpassed that of, the normal, uninjured, side of the diaphragm, as well as the diaphragm of uninjured animals (Fig. 2b, c). Although occasionally the duration of inspiratory bursting could approach normal levels, it remained shorter on average than that of the uninjured side (Fig. 2a, b).

Although recording from the diaphragm can give semiquantitative indications of muscle activity and neuromuscular junction effectiveness, phrenic nerve recordings under standardized conditions allow more-quantitative measurements of the total motor output of the phrenic nucleus. When recording phrenic nerve activity at a chronic stage after C2 hemisection and treatment, the peak amplitude followed trends similar to those revealed by EMG analyses, with ChABC-treated, grafted animals having the highest values (Supplementary Fig. 6). Although peak amplitude was high, the reduction in burst duration for both EMG and neurogram activity suggests



**Figure 2 | A PNG enhances diaphragmatic EMG activity.** After C2 hemisection and implantation of a PNG, animals that received ChABC treatment showed an increase in EMG activity in the diaphragm over time to almost normal levels. **a**, For 6 or more weeks after C2 hemisection, there was minimal activity in the hemidiaphragm ipsilateral to the lesion in animals that received ChABC treatment and an autologous PNG. At 12 weeks, there was substantial recovery of hemidiaphragmatic inspiratory activity. **b**, At 12 weeks, animals that had received a PNG and ChABC treatment had EMG activity close to that of an uninjured animal and better than that of grafted animals that received only saline treatment. **c**, At 12 weeks after C2 hemisection, more animals that received a PNG and ChABC treatment had recovered than those that received the alternative treatments (top left). A semi-quantitative analysis showed that there was no difference in the frequency of breaths between the groups (top right). C2-hemisected animals with a PNG and ChABC treatment had a higher average peak amplitude for the raw inspiratory bursts than animals in the other groups (bottom left). Although burst duration could, on occasion, reach almost normal levels (as shown in **a** and **b**), on balance there was no difference in the average duration of inspiratory bursts between animals that showed recovery (bottom right). The control (bottom panels) is the inspiratory peak amplitude (left) or burst duration (right) of the hemidiaphragm contralateral to the lesion. Darker coloured bars represent all animals in a group, whereas lighter bars represent only the animals that showed recovered activity. \*\*, significantly different compared with saline-treated animals and ChABC-treated animals;  $P < 0.02$ . \*, significantly different compared with grafted animals treated with saline,  $P < 0.05$ . Error bars, s.e.m.

either that an increased phrenic motor neuron threshold develops after injury and/or regeneration or that the number of reinnervated motor neurons is suboptimal. Inducing the expression of neurotrophins to lure regenerating axons towards the phrenic motor pool, or transiently diminishing the expression of the *PTEN* gene, might further augment connectivity and correct this deficit<sup>21–23</sup>. Interestingly, at 6 and 9 weeks after C2 hemisection and grafting, the emerging hemidiaphragmatic activity, although mostly patterned with the contralateral side, was neither uniform nor consistent. Instead, specific motor unit activity occurred intermittently, with spikes of differing

amplitudes appearing and disappearing throughout the recording session (Supplementary Fig. 7). In preliminary studies of C2 hemilectomized animals at 6 months after the various treatments (saline, ChABC, PNG and saline, and PNG and ChABC;  $n = 3$  per group), the recovery in all groups remained at the level equivalent to that at 3 months (data not shown).

### Anatomical evidence of regeneration

Immunohistochemistry and tracing experiments at the 12-week time point suggested robust regeneration into the PNG. An abundance of tau, a marker for microtubule-associated proteins normally found in axons, indicated that there was significant regeneration of axons into the pre-degenerated graft (Fig. 3a, b). Similarly to a normal dorsal-root entry zone, at the interface between the central nervous system (CNS) and the ChABC-treated graft, glial fibrillary acidic protein (GFAP)-positive astrocytes extended a short distance into the graft and were aligned with the regenerated axons<sup>24</sup> (Fig. 3a, b). However,

in saline-treated animals, GFAP-positive astrocytes were haphazardly organized in a geometry that could obstruct the exit of regenerated axons at the interface (Fig. 3a, b).

At the distal interface between the graft and the spinal cord in ChABC-treated, grafted animals, there was a suggestion of deep penetration of both tau-positive and 5-HT-positive axons into CNS tissue (the CNS compartment is demarcated by the presence of GFAP) (Fig. 3a–d). In a group of animals in which the tracers dextran amine Texas Red or biotinylated dextran amine (BDA) were injected into the medulla to label bulbospinal tracts, reconstructions of multiple sections confirmed that there was substantial regeneration into the graft and further into the cervical spinal cord (Fig. 3e). The regenerated axons tended to remain within their segment of re-entry. Furthermore, labelled axons were associated with synapsin puncta, suggesting the possibility of functional reconnection (Fig. 3f). At 12 weeks, quantification of 5-HT immunoreactivity in the ventral horns near C4 motor neurons showed significantly more serotonergic fibres in ChABC-treated, grafted animals than in saline-treated, grafted animals (Supplementary Fig. 8). Interestingly, even though the extent of recovery between ChABC only and ChABC-treated, grafted animals differed markedly, the 5-HT-positive fibre densities in the vicinity of the phrenic motor pools were comparable (Fig. 2c and Supplementary Fig. 8).

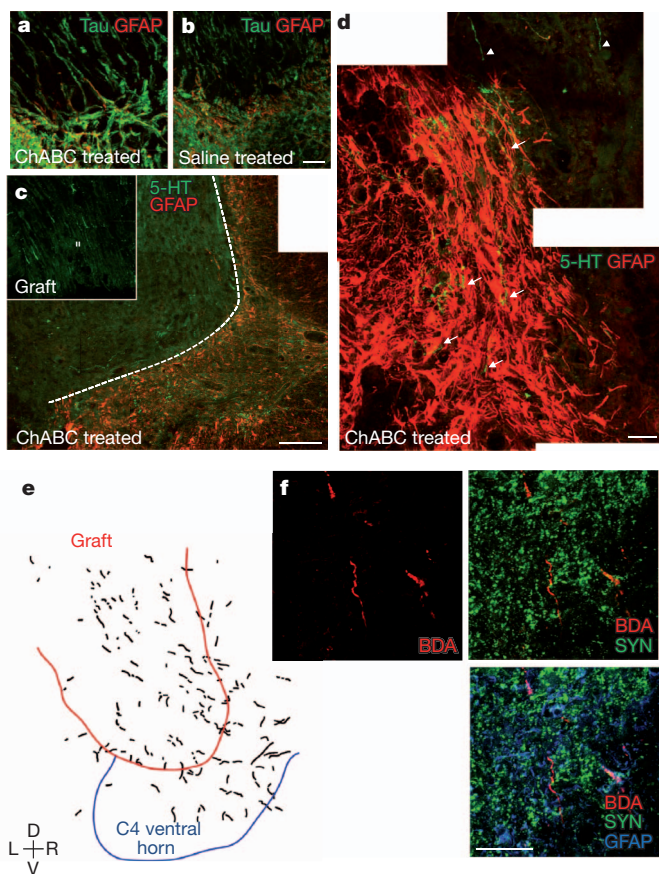
Taken together, these results suggest that the raphe-spinal system is not acting alone in fostering restoration of hemidiaphragmatic function. When dextran amine Texas Red was injected directly into the graft, to allow a glimpse of the entire population of neurons that projected axons through the graft, the results indicated that there was a small portion of retrogradely labelled neurons within the raphe nuclei and, more importantly, the rVRG, a nucleus that is crucial for caudally projecting respiratory rhythm and drive (Supplementary Fig. 9). However, there were also a significant number of projections from the medial medullary reticular formation (Supplementary Fig. 9). Additionally, there were a few labelled cell bodies in the dorsolateral medulla (Supplementary Fig. 9).

### Consequences of transecting the graft

In animals with restored hemidiaphragmatic function, transection of the graft led to complete elimination of inspiratory hemidiaphragmatic activity ipsilaterally (along with a compensatory increase in frequency and amplitude on the contralateral side), although this occurred in a manner distinctly different from that following acute hemisection in an intact spinal cord (see below). The marked reduction of EMG activity when the bridge was transected and the compensatory changes in the contralateral side strongly suggest that recovery of diaphragm motor function was primarily mediated through regeneration of respiratory-related axons and had a crucial role in ventilation of the animal (Fig. 4a, b).

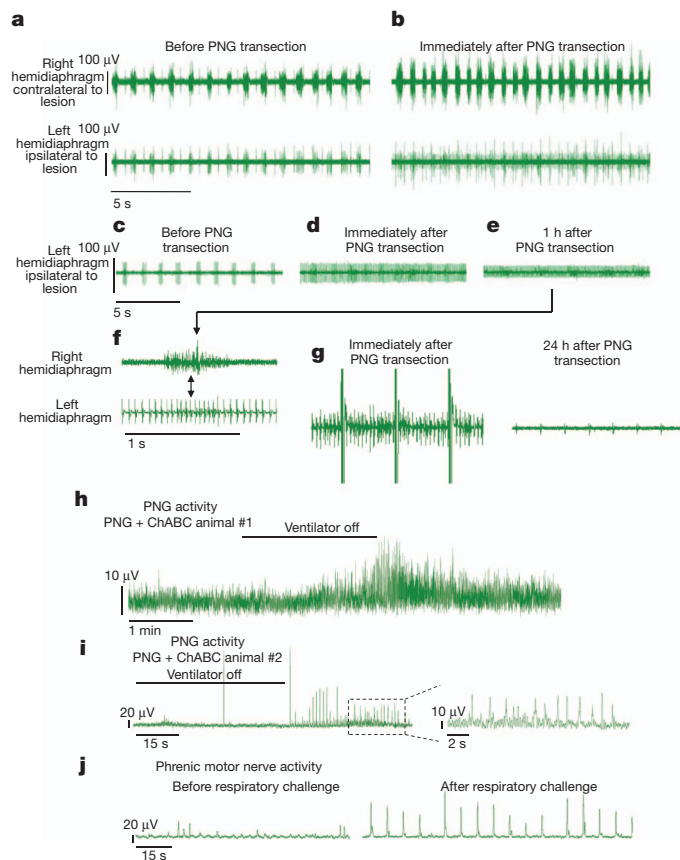
Surprisingly, in addition to eventually abolishing inspiratory activity, transection of the PNG initially led to an unusual increase in overall tonic EMG activity (Fig. 4g), which varied in frequency between animals and which does not usually occur during or after a C2 hemisection (Supplementary Fig. 10). Thus, tonic activity in the motor neurons themselves results only when the regenerated fibres reinnervating the spinal cord are eliminated. The pattern of this activity was reminiscent of extracellular recordings of interneurons that are normally found deep within the spinal respiratory circuitry: during times of inspiration (identified by breaths on the contralateral side), there were repeated, transient increases in the tonic spiking frequency<sup>25</sup> (Fig. 4b, d, f). At acute stages after graft lesion, any remnants of patterned inspiratory activity buried within the tonic bursting episode could be attributed to these interneurons that, in turn, might receive inputs from the crossed phrenic pathway<sup>11,26</sup>. The increased tonic activity after graft lesion subsided at 1 h and was absent when assessed at 24 h (Fig. 4e, g).

These results suggest that, after spinal cord injury, regenerated axons may incorporate the activity of interneurons and/or propriospinal



**Figure 3 | ChABC treatment promotes axon regeneration.** There is significant regeneration of axon fibres in the PNG and back into the CNS after ChABC treatment. **a, b**, In animals that received ChABC, at the graft–spinal cord interface, astroglial processes from the spinal cord, identified by GFAP labelling (red), have aligned with tau-positive axons (green) that have regenerated back into the CNS. In saline-treated animals, it appears that the astrocytes form a barrier-like structure at the interface. Scale bar, 40  $\mu$ m. **c**, In ChABC-treated animals, only a small proportion of the regenerated axons in the graft are serotonergic, with the dashed line demarcating the graft–spinal cord interface (green; top left). Scale bar, 200  $\mu$ m. **c, d**, In ChABC-treated animals, serotonergic fibres (arrows) penetrated deep into the CNS (identified by GFAP, red) from the graft (arrowheads). Scale bar, 40  $\mu$ m (**d**). **e**, Anterograde tracing from the medulla with dextran amine Texas Red showed that fibres had regenerated in the graft and back into the grey matter of the spinal cord (C4 ventral horn) in ChABC-treated animals. D, dorsal; L, left; R, right; V, ventral. **f**, BDA labelling and immunohistochemistry show the close proximity of the regenerated fibres with synapsin (SYN) puncta in ChABC-treated animals. Scale bar, 40  $\mu$ m.





**Figure 4 | PNG transection initially increases tonic diaphragmatic EMG activity then completely eliminates activity.** Transection of the PNG after recovery led to increased tonic EMG activity of the diaphragm. **a**, At 12 weeks after C2 hemisection in an animal that had received a graft and ChABC treatment, there was recovery ipsilateral to the lesion that was rhythmic and synchronous with the contralateral side. **b**, Immediately after transection of the graft, there was a reduction in the recovered activity but an increase in tonic activity and compensatory changes on the contralateral side. **c–e**, In a second recovered animal in which the PNG was transected, the tonic activity declined after 1 h. **f**, In this animal, at 1 h after PNG transection, during times of inspiration (indicated by the upper trace, of the right hemidiaphragm), there was an increase in the spiking frequency in the resultant tonic activity of the left hemidiaphragm. **g**, In a third animal, the tonic activity was absent at 24 h after graft transection. **h, i**, In two animals that had received a graft and ChABC treatment, there was an abundance of activity in the graft that could be augmented during and after respiratory challenge, suggesting regeneration of respiratory-related axons. **j**, Phrenic motor nerve activity after a similar challenge mirrored the pattern of increased activity in the graft in **h** and **i**.

neurons that innervate motor neurons, leading to proper firing and restoration of function. Other potential mechanisms that underlie this observation include a disinhibition of phrenic motor neurons that results directly from removing the regenerated supraspinal inputs or alterations in a phenomenon known as homeostatic plasticity, in which neurons in denervated tissues can reacquire stable, repetitive firing characteristics<sup>27</sup>.

When recording from the implanted graft itself, although there was a continuous barrage of firing during full artificial ventilation, there was a strong and persistent increase in activity in response to respiratory challenge (that is, when the ventilator was temporarily shut down) (Fig. 4h, i). Such enhanced activity clearly suggests that among the many axons in the graft, there was also a contingent of respiratory-related axons, confirming earlier work<sup>15,17</sup>. Furthermore, and rather remarkably, the regular firing pattern of the subgroup of axons in the graft with augmented activity closely mirrored the pattern of enhanced activity that developed in the ipsilateral phrenic nerve (Fig. 4j).

Importantly, restored output activity in the phrenic nerve was never contaminated with non-respiratory-related input activity that was present in the graft (Fig. 4j).

## Discussion

Taken together, our experiments show that in adult rats, robust activation of a critical muscle that is paralysed by spinal cord injury can be returned through long distance regeneration of axons into the vicinity of motor neurons in the denervated grey matter. EMG recordings showed that recovery takes place very slowly but is remarkably strong when it finally manifests. Importantly, ChABC treatment on its own, although sufficient to initiate rapid functional recovery in most C2-hemisected animals, was ultimately limited. Perhaps inducing respiratory drive through intermittent hypoxia, theophylline administration, or channel rhodopsin 2 transfection in the C4 ventral horn plus light stimulation each in combination with ChABC would result in a more rapid or higher level of recovery without the need for grafting or even following grafting<sup>10,28–31</sup>.

In addition to returning function to a paralysed muscle, and the clinical implications of restoring the ability to breathe after spinal cord injury, our results also provide new insight into the role of adult CNS plasticity and reorganization after injury and regeneration. Indeed, the increased tonic activity of the diaphragm that results from transecting the PNG suggests that widespread alterations in spinal circuitry occur after injury. In particular, our results suggest an important role for interneurons, the activity of which can be recruited by regenerating axons. However, it has not yet been proven that such interneurons are, in fact, the anatomical substrate for recovery. It is also possible that they are extraneous and being inhibited, at least when the graft is intact. In addition, the adult CNS, even after injury, has the remarkable ability to organize a composite of regenerating axons and their diverse signals into meaningful synaptic connections, as well as eliminate or silence abnormal and potentially debilitating connections that could result in misfiring muscle activity. Perhaps these endogenous mechanisms in the CNS determine the time course of recovery. These processes could include the pruning or inhibition of unwanted (non-respiratory-related) synaptic connections, the remyelination of necessary (as well as unnecessary) regenerated fibres, or the competition with spared tracts for a finite number of postsynaptic sites. Undoubtedly, these findings provide many new avenues for research into CNS regeneration and promoting recovery after spinal cord injury.

## METHODS SUMMARY

Sprague Dawley female rats (240–300 g) received a C2 hemisection and at the same time either saline or ChABC at the level of the phrenic nucleus, ipsilateral to the lesion. Some animals received a pre-degenerated, autologous peripheral tibial nerve bridge, and in these animals, a single injection of ChABC was administered at the C2 lesion site where the proximal end of the graft was inserted, to enhance axon entry. After 1 week, the spinal cord was re-exposed, and the distal end of the graft was inserted into a pocket at C4 together with a single ChABC injection.

For the EMG recordings, two bipolar electrodes, connected to amplifiers and a data acquisition system, were inserted into the left and right hemidiaphragms. For phrenic neurogram recordings, animals were vagotomized and ventilated, and their femoral blood vessels were cannulated to monitor blood pressure and administer drugs. The left phrenic nerve, ipsilateral to the lesion, was isolated, transected, desheathed, placed on bipolar silver electrodes and covered with mineral oil, and its activity was recorded under standardized conditions. The preparation of the animal to record from the graft itself was similar to the above procedures; however, the spinal cord was re-exposed, and the graft isolated. The freed graft was then placed on the electrodes and its activity recorded.

For the immunohistochemistry experiments, animals were perfused with PBS and 4% paraformaldehyde. The medulla and spinal cord were collected and sectioned on a cryostat at a 20-μm thickness and were then processed for detection of the relevant molecules. For the anatomical tracing studies, BDA and/or dextran amine Texas Red was injected into one of the following: the diaphragm, to label phrenic motor neurons in a retrograde manner; the graft, to label the population of regenerating medullary cells in a retrograde manner; or the medulla, to label the regenerated axons in the graft and back into the spinal cord.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 20 July 2010; accepted 13 May 2011.**

- Davies, S. J. *et al.* Regeneration of adult axons in white matter tracts of the central nervous system. *Nature* **390**, 680–683 (1997).
- Bradbury, E. J. *et al.* Chondroitinase ABC promotes functional recovery after spinal cord injury. *Nature* **416**, 636–640 (2002).
- Silver, J. & Miller, J. H. Regeneration beyond the glial scar. *Nature Rev. Neurosci.* **5**, 146–156 (2004).
- Massey, J. M. *et al.* Chondroitinase ABC digestion of the perineuronal net promotes functional collateral sprouting in the cuneate nucleus after cervical spinal cord injury. *J. Neurosci.* **26**, 4406–4414 (2006).
- García-Alias, G., Barkhuysen, S., Buckle, M. & Fawcett, J. W. Chondroitinase ABC treatment opens a window of opportunity for task-specific rehabilitation. *Nature Neurosci.* **12**, 1145–1151 (2009).
- Hunanyan, A. S. *et al.* Role of chondroitin sulfate proteoglycans in axonal conduction in mammalian spinal cord. *J. Neurosci.* **30**, 7761–7769 (2010).
- Schachtrup, C. *et al.* Fibrinogen triggers astrocyte scar formation by promoting the availability of active TGF- $\beta$  after vascular damage. *J. Neurosci.* **30**, 5843–5854 (2010).
- Choi, H. *et al.* Respiratory abnormalities resulting from midcervical spinal cord injury and their reversal by serotonin 1A agonists in conscious rats. *J. Neurosci.* **25**, 4550–4559 (2005).
- Fuller, D. D. *et al.* Cervical spinal cord injury upregulates ventral spinal 5-HT<sub>2A</sub> receptors. *J. Neurotrauma* **22**, 203–213 (2005).
- Golder, F. J. & Mitchell, G. S. Spinal synaptic enhancement with acute intermittent hypoxia improves respiratory function after chronic cervical spinal cord injury. *J. Neurosci.* **25**, 2925–2932 (2005).
- Goshgarian, H. G. The crossed phrenic phenomenon: a model for plasticity in the respiratory pathways following spinal cord injury. *J. Appl. Physiol.* **94**, 795–810 (2003).
- Hadley, S. D., Walker, P. D. & Goshgarian, H. G. Effects of the serotonin synthesis inhibitor *p*-CPA on the expression of the crossed phrenic phenomenon 4 h following C2 spinal cord hemisection. *Exp. Neurol.* **160**, 479–488 (1999).
- Hodges, M. R. *et al.* Defects in breathing and thermoregulation in mice with near-complete absence of central serotonin neurons. *J. Neurosci.* **28**, 2495–2505 (2008).
- David, S. & Aguayo, A. J. Axonal elongation into peripheral nervous system ‘bridges’ after central nervous system injury in adult rats. *Science* **214**, 931–933 (1981).
- Munz, M. *et al.* Functional activity of rat brainstem neurons regenerating axons along peripheral nerve grafts. *Brain Res.* **340**, 115–125 (1985).
- Gauthier, P. & Rasminsky, M. Activity of medullary respiratory neurons regenerating axons into peripheral nerve grafts in the adult rat. *Brain Res.* **438**, 225–236 (1988).
- Gauthier, P. *et al.* Functional reconnections established by central respiratory neurons regenerating axons into a nerve graft bridging the respiratory centers to the cervical spinal cord. *J. Neurosci. Res.* **70**, 65–81 (2002).
- Fouad, K. *et al.* Combining Schwann cell bridges and olfactory-ensheathing glia grafts with chondroitinase promotes locomotor recovery after complete transection of the spinal cord. *J. Neurosci.* **25**, 1169–1178 (2005).
- Houle, J. D. *et al.* Combining an autologous peripheral nervous system ‘bridge’ and matrix modification by chondroitinase allows robust, functional regeneration beyond a hemisection lesion of the adult rat spinal cord. *J. Neurosci.* **26**, 7405–7415 (2006).
- Yang, L. J. *et al.* Sialidase enhances spinal axon outgrowth *in vivo*. *Proc. Natl Acad. Sci. USA* **103**, 11057–11062 (2006).
- Zhou, L., Baumgartner, B. J., Hill-Felberg, S. J., McGowen, L. R. & Shine, H. D. Neurotrophin-3 expressed *in situ* induces axonal plasticity in the adult injured spinal cord. *J. Neurosci.* **23**, 1424–1431 (2003).
- Alto, L. T. *et al.* Chemotropic guidance facilitates axonal regeneration and synapse formation after spinal cord injury. *Nature Neurosci.* **12**, 1106–1113 (2009).
- Liu, K. *et al.* PTEN deletion enhances the regenerative ability of adult corticospinal neurons. *Nature Neurosci.* **13**, 1075–1081 (2010).
- McPhail, L. T., Plunet, W. T., Das, P. & Ramer, M. S. The astrocytic barrier to axonal regeneration at the dorsal root entry zone is induced by rhizotomy. *Eur. J. Neurosci.* **21**, 267–270 (2005).
- Hayashi, F., Hinrichsen, C. F. & McCrimmon, D. R. Short-term plasticity of descending synaptic input to phrenic motoneurons in rats. *J. Appl. Physiol.* **94**, 1421–1430 (2003).
- Lane, M. A. *et al.* Cervical prephrenic interneurons in the normal and lesioned spinal cord of the adult rat. *J. Comp. Neurol.* **511**, 692–709 (2008).
- Turrigiano, G. G. & Nelson, S. B. Homeostatic plasticity in the developing nervous system. *Nature Rev. Neurosci.* **5**, 97–107 (2004).
- Nantwi, K. D., El-Bohy, A. & Goshgarian, H. G. Actions of systemic theophylline on hemidiaphragmatic recovery in rats following cervical spinal cord hemisection. *Exp. Neurol.* **140**, 53–59 (1996).
- Fuller, D. D., Johnson, S. M., Olson, E. B. Jr & Mitchell, G. S. Synaptic pathways to phrenic motoneurons are enhanced by chronic intermittent hypoxia after cervical spinal cord injury. *J. Neurosci.* **23**, 2993–3000 (2003).
- Teng, Y. D. *et al.* Serotonin 1A receptor agonists reverse respiratory abnormalities in spinal cord-injured rats. *J. Neurosci.* **23**, 4182–4189 (2003).
- Alilain, W. J. *et al.* Light-induced rescue of breathing after spinal cord injury. *J. Neurosci.* **28**, 11862–11870 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This project was funded by The Christopher and Dana Reeve Foundation (W.J.A.), the International Spinal Research Trust (W.J.A.), the National Institute of Neurological Disorders and Stroke, grants NS25713 and NS060767 (J.S.), and the National Heart, Lung and Blood Institute, grant HL080318 (T.E.D.). We offer special thanks to the Brumagin Memorial Fund (J.S.) and the Ellen Becker Neuroscience Regenerative Medicine Research Fund (J.S.). We also thank J. D. Houle and V. J. Tom for teaching us the peripheral nerve grafting technique.

**Author Contributions** W.J.A. performed the surgeries, recorded and analysed the electrophysiological data, and completed the immunocytochemical, tracing and histological experiments. K.P.H. helped with the immunocytochemical detection and quantification. H.H. assisted with the animal care, surgeries and data quantification, as well as performed all of the tissue processing. T.E.D. gave technical guidance and helped discuss the results. W.J.A. and J.S. designed all of the studies, analysed the data and wrote the paper. All authors discussed the data and helped to prepare the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.S. (jxs10@case.edu) or W.J.A. (wja4@case.edu).



## METHODS

**C2 hemisection and grafting.** Adult female Sprague Dawley rats (retired breeders, 240–300 g) were anaesthetized with a ketamine ( $70\text{ mg kg}^{-1}$  body weight) and xylazine ( $7\text{ mg kg}^{-1}$  body weight) solution administered intraperitoneally. After the animals reached a surgical plane of anaesthesia, they were prepared for surgery by shaving the dorsal surface of the neck area and scrubbing with Betadine and 70% ethyl alcohol. A midline incision of approximately 4 cm was made, and the paravertebral muscles were retracted. A multilevel laminectomy was performed to expose the upper cervical segments of the spinal cord. The dura was then cut. Using a microblade, a hemisection was made on the animal's left spinal cord, caudal to the C2 dorsal roots and starting at the midline and extending to the lateral most extent of the spinal cord. Sham hemisected animals received all procedures but the lesion.

Animals that received a PNG were anaesthetized as above, and the left tibial nerve was uncovered and transected 1 week before the C2 hemisection, for pre-degeneration. At the time of C2 hemisection, the tibial nerve was dissected out (1–1.5 cm) and placed in the C2 hemisection along with ChABC. After a further 1 week, the spinal cord was again revealed, and the distal end of the graft was placed into a slit made at the C4 spinal cord with ChABC. At both proximal and distal ends of the graft, the epineurium and dura were sutured together, for stability, with 9-0 polypropylene suture. To see a visual demonstration of the grafting technique, see ref. 32.

**ChABC injections.** For animals that received either ChABC ( $20\text{ U ml}^{-1}$ ) only or a saline vehicle control only, a pulled pipette attached to a Nanoject II (Drummond Scientific Company) was stereotactically placed 1.1 mm to the left of the spinal cord midline and 1.6 mm ventral from the dorsal surface of the spinal cord, in close proximity to the phrenic nucleus. After placement, approximately 250 nl drug or vehicle was administered per injection. For each 250-nl injection, the ChABC spread effect was approximately  $20.5\text{ mm}^3$ . Three injections were made at  $\sim 1\text{ mm}$  apart at the C3 to C5 spinal cord.

In animals that received a PNG, 500 nl ChABC ( $20\text{ U ml}^{-1}$ ) was injected into the C2 lesion area with the Nanoject before implantation of the proximal end. After 1 week, 250 nl ChABC ( $20\text{ U ml}^{-1}$ ) was injected into a slit made tangentially to the plane of the spinal cord at a depth of about 1 mm at the C4 cervical level before insertion of the distal end.

After this, the muscle layers were drawn back together with 3-0 vicryl, and the skin was stapled together with wound clips. The animals received Marcaine and Buprenorphine for analgesic purposes. After surgery and if animals appeared to be dehydrated, saline (5–10 ml) was administered subcutaneously. All animals were housed in groups of two to three and exposed to a normal dark–light cycle with free access to food (normal rodent chow) and water. All animal procedures were approved by Case Western Reserve University's Institutional Animal Care and Use Committee.

**Physiological recordings.** For diaphragmatic EMG recordings, animals were anaesthetized with ketamine and xylazine, and an 8-cm incision was made at the base of the rib cage to expose the abdominal surface of the diaphragm. Bipolar electrodes, connected to an amplifier and data acquisition system (CED1401 with Spike2 software, Cambridge Electronic Design), were inserted bilaterally into the crural part of the diaphragm, and EMG muscle activity was recorded. The EMG signal was band-pass filtered between 30 and 3,000 Hz (P511 amplifier, Grass Technologies). If animals were to survive, the abdominal muscles were sutured together with 3-0 vicryl, and the skin was closed with wound clips. For induction of the crossed phrenic pathway, the above procedure was performed after the phrenic nerve was isolated, contralateral to the lesion (from a ventral approach), and then transected. For transection of the graft during EMG recording, the spinal cord was re-exposed, and the graft was isolated and severed. The graft in the living animal is not tightly apposed to the dura and can still be manipulated, lifted and cut without touching or damaging the dorsal surface of the spinal cord.

For phrenic nerve recordings, the animals were first anaesthetized with urethane ( $1.6\text{ g kg}^{-1}$  body weight, administered intraperitoneally). The femoral vein and artery were then cannulated with PE 50 tubing, to administer fluids and drugs and to monitor blood pressure. PE-240 tubing connected to a ventilator (Harvard Apparatus) was inserted into an opening made in the trachea to ventilate the animal with a 1:1 mixture of oxygen to air. Both vagus nerves were transected to abolish mechanoreceptor feedback and entrainment. During the whole recording procedure, the animal was paralysed with vancuronium; the end-tidal  $\text{CO}_2$  levels

were monitored (small animal  $\text{CO}_2$  monitor, Kent Scientific); and the animal was placed on a heated circulating-water pad. From the dorsal approach, the left phrenic nerve was dissected, transected at a point far distal from the CNS and desheathed. The dissected nerve was then placed on bipolar silver electrodes and covered with mineral oil. Under standardized recording conditions, phrenic nerve activity was amplified ( $20,000\times$ ) and band-pass filtered (30–1,000 Hz) (QP511 amplifier system, Grass Technologies), and activity was monitored both visually and through an audio monitor. The signal was recorded and analysed using the CED1401/Spike2 data acquisition system. For recording the activity of the surgically applied PNG, the above procedures were performed; however, instead of dissecting the phrenic nerve, the spinal cord was re-exposed, and the PNG was isolated. The proximally attached graft was then placed on the silver bipolar electrodes, and its activity was recorded.

**Quantification and statistical analysis.** During recording, raw diaphragmatic EMG signal or phrenic nerve activity was rectified and integrated using Spike2 software. Frequency was determined by counting the total number of inspiratory bursts for 5 min. Peak amplitude and burst duration of inspiratory bursts were measured through Spike2, for at least ten breaths per animal for each group. Peak amplitude values were standardized to either a 10 or 100  $\mu\text{V}$  calibration pulse and contralateral hemidiaphragm or homolateral phrenic nerve activity. Values were averaged, and statistical analysis was performed using a one-way analysis of variance (ANOVA) and Tukey's post hoc analysis, with Minitab 15 software. A  $P$  value less than 0.05 was considered significant. All error bars are s.e.m.

**Immunohistochemistry and tracing.** At the time of perfusion, animals were anaesthetized with ketamine and xylazine and perfused first with 50 ml PBS and then with 250 ml 4% paraformaldehyde in PBS. The spinal cord was removed and fixed in 4% paraformaldehyde overnight and then cryoprotected in 30% sucrose in PBS until sectioning. Immediately before sectioning, a pinhole was made on the side contralateral to the lesion to denote laterality. The spinal cord was sectioned on a cryostat (Hacker Instruments & Industries) at a thickness of 20  $\mu\text{m}$  and mounted on SuperFrost coated slides.

Mounted sections were washed three times with PBS and then blocked in 5% normal goat serum and 0.1% BSA in PBS. Triton X-100 (0.1%) was added to the blocking buffer depending on the antigen being studied. Sections were then incubated in primary antibody diluted in blocking buffer for two nights at  $4^\circ\text{C}$ . The primary antibodies used were mouse anti-chondroitin sulphate (1:200, Sigma), 2B6 (1:200, Sigma), anti-NeuN (1:500, Chemicon), anti-synapsin (1:1,000, Chemicon) and anti-GFAP (1:500, Sigma) antibodies, and rabbit anti-5-HT (1:15,000, ImmunoStar) and anti-tau (1:1,500, Abcam) antibodies. PNN and glycosaminoglycan chains were detected with *Wisteria floribunda* (WFA) lectin conjugated to biotin (1:50, Sigma). The next day, the sections were washed extensively with PBS and incubated in the appropriate secondary antibody or avidin substrate conjugated to Alexa Fluor 488, 594 or 633 (1:500, Molecular Probes) for two days. After extensive washing, the sections were mounted, cover-slipped and viewed with a confocal microscope (Zeiss). Pixel intensity was measured for images taken using a standard fluorescent microscope (Leica) with a standard exposure setting and was analysed using MetaMorph Microscopy Automation & Image Analysis Software (Molecular Devices).

For all tracing studies, tracers were injected 1 week before animal perfusion. To retrogradely label phrenic motor neurons, 10  $\mu\text{l}$  0.4% 3,000-Da dextran amine Texas Red (Molecular Probes) in PBS was injected five times into the hemidiaphragm, ipsilateral to the hemisection, with a Hamilton syringe. To label regenerating medullary axons in the graft and back into the spinal cord, the medulla was exposed, and 500 nl 10% 10,000-Da dextran amine Texas Red or BDA (Molecular Probes) was injected bilaterally, at  $\sim 2.5$ –3 mm lateral to the obex by using a Nanoject. Labelled axons from five sections were traced onto a composite, by using a light box. To retrogradely label the medullary cell bodies projecting axons into the graft, the spinal cord and graft area was re-exposed, and 125 nl 0.4% 3,000-Da dextran amine Texas Red was injected into the graft with a Nanoject. To identify the rostrocaudal position of the sections, the transition from the central canal to the fourth ventricle was identified, and the labelled cell bodies were positioned onto the coronal sections of a rat brain atlas<sup>33</sup>.

32. Houle, J. D. *et al.* Combining peripheral nerve grafting and matrix modulation to repair the injured rat spinal cord. *J. Vis. Exp.* doi:10.3791/1324 (2009).

33. Paxinos, G. & Watson, C. *The Rat Brain in Stereotaxic Coordinates* (Academic, 1998).

# Dicer recognizes the 5' end of RNA for efficient and accurate processing

Jong-Eun Park<sup>1\*</sup>, Inha Heo<sup>1\*</sup>, Yuan Tian<sup>2</sup>, Dharendra K. Simanshu<sup>2</sup>, Hyeshik Chang<sup>1</sup>, David Jee<sup>1</sup>, Dinshaw J. Patel<sup>2</sup> & V. Narry Kim<sup>1</sup>

**A hallmark of RNA silencing is a class of approximately 22-nucleotide RNAs that are processed from double-stranded RNA precursors by Dicer. Accurate processing by Dicer is crucial for the functionality of microRNAs (miRNAs). The current model posits that Dicer selects cleavage sites by measuring a set distance from the 3' overhang of the double-stranded RNA terminus. Here we report that human Dicer anchors not only the 3' end but also the 5' end, with the cleavage site determined mainly by the distance (~22 nucleotides) from the 5' end (5' counting rule). This cleavage requires a 5'-terminal phosphate group. Further, we identify a novel basic motif (5' pocket) in human Dicer that recognizes the 5'-phosphorylated end. The 5' counting rule and the 5' anchoring residues are conserved in *Drosophila* Dicer-1, but not in *Giardia* Dicer. Mutations in the 5' pocket reduce processing efficiency and alter cleavage sites *in vitro*. Consistently, miRNA biogenesis is perturbed *in vivo* when Dicer-null embryonic stem cells are replenished with the 5'-pocket mutant. Thus, 5'-end recognition by Dicer is important for precise and effective biogenesis of miRNAs. Insights from this study should also afford practical benefits to the design of small hairpin RNAs.**

RNAse III proteins have central roles in RNA silencing by processing double-stranded (ds)RNA precursors into small RNA duplexes<sup>1</sup>. Drosha cleaves a primary precursor of miRNA (pri-miRNA) to release a hairpin-shaped pre-miRNA<sup>2</sup>. Dicer cuts the pre-miRNA near the terminal loop and generates a short miRNA duplex<sup>3–7</sup>. Dicer also participates in small interfering RNA (siRNA) production from long RNA duplexes. One strand of the small RNA duplex is subsequently loaded onto the Argonaute protein to yield an active RNA-induced silencing complex<sup>8–10</sup>.

Precise selection of cleavage sites by RNAse III enzymes is critical in miRNA biogenesis because alterations in the cleavage site can change the abundance and/or targeting specificity of the miRNA. To determine the cleavage site, Drosha and Dicer recognize certain RNA structures and cleave a fixed distance away from the structure. In the case of Drosha, its cofactor DGCR8 (also known as Pasha) binds to the base of the stem-loop structure and locates the catalytic site of Drosha ~11 base pairs (bp) away from the single-stranded (ss)RNA–dsRNA junction<sup>11</sup>. Thus, the ssRNA–dsRNA junction serves as the reference point for Drosha processing.

Dicer, on the other hand, is known to measure ~22 nucleotides away from the 3' end of the open terminus of dsRNA helices<sup>12–14</sup>. The crystal structure of Dicer from *Giardia intestinalis* in its free state and biochemical analyses indicated that the PAZ domain of Dicer anchors the 3' overhang of the dsRNA terminus, and that the dsRNA stem is placed along the positively charged protein extension to reach the catalytic centre of Dicer<sup>15,16</sup>. This spatial arrangement would enable Dicer to measure a fixed distance from the 3' end of the terminus (the '3' counting model').

Recent studies have shown that certain pre-miRNAs are modified at their 3' end in the cell. The most common type of pre-miRNA modification is addition of non-templated uridyl residues<sup>17–20</sup>. According to the current 3' counting model, such 3'-end modifications are expected to shift the Dicer cleavage site towards the open terminus. This would change the seed sequences of miRNAs originating from the 3' strand and might even alter strand selection<sup>21,22</sup>.

## Human Dicer counts from the 5'-phosphorylated end

To understand the impact of pre-miRNA uridylation on Dicer processing, we prepared synthetic pre-let-7a-1 with extra uridine residues at the 3' end (Fig. 1a, left). The RNA was labelled at the 5' end with [ $\gamma$ -<sup>32</sup>P] ATP and incubated with immunopurified human Dicer. With the 3'-elongated substrates, we expected to observe a shift of the cleavage site, which would yield shorter products from the 5' strand. Surprisingly, the size of the major cleavage products remained the same (22 nucleotides) (purple arrowheads), indicating that the pre-let-7a-1 variants were cleaved at the same site regardless of the 3' extension. We observed similar cleavage patterns when pre-miR-16-1 variants were used (Supplementary Fig. 1a). Substrates labelled at the 3' ends were also cleaved at the same sites, excluding the possibility that the 3' extension was trimmed back by a contaminating nuclease (Supplementary Fig. 1b). This processing pattern was not influenced by the sequences of the nucleotides added to the 3' overhang: addition of adenosine or cytidine instead of uridine gave comparable results (Supplementary Fig. 1a and data not shown).

We next examined duplex RNAs with varying 3' overhangs (Fig. 1b, left). Like pre-miRNAs, the predominant products from the dsRNAs were 22 nucleotides in length in spite of the differences at the 3' overhangs, indicating that human Dicer may not be dependent on the 3' end for cleavage site selection. We noticed another group of minor products, which were shortened as the 3' overhang was elongated (green arrowheads). This minor cleavage pattern is expected of the 3' counting model. Small amounts of 3' counting products were also generated from pre-miRNAs (Fig. 1a and Supplementary Fig. 1a).

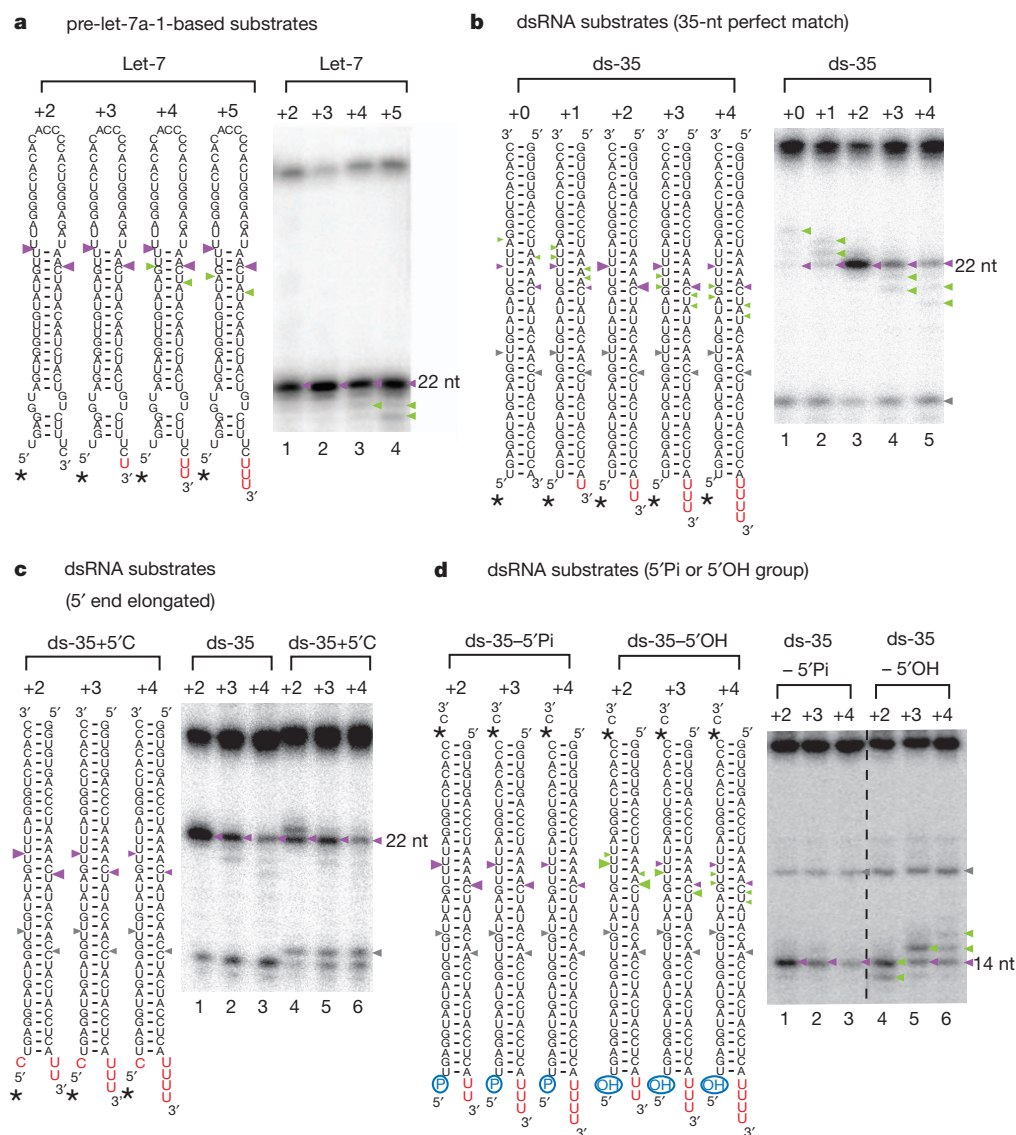
To summarize, we observed two types of Dicer cleavage events occurring in parallel. In the type predominant for the substrates used here, the cleavage site does not change upon 3' end elongation. In another type, the cleavage site is determined based on the distance from the 3' end. These results indicate that in addition to the 3' end, Dicer may recognize other structural feature(s) of the RNA substrate.

Substrates with a 2-nucleotide 3' overhang were cleaved most uniformly and efficiently, indicating that Dicer binds to these canonical

<sup>1</sup>School of Biological Sciences, Seoul National University, Seoul 151-742, Korea. <sup>2</sup>Structural Biology Program, Memorial-Sloan Kettering Cancer Center, New York, New York 10065, USA.

\*These authors contributed equally to this work.





**Figure 1 | Human Dicer counts 22 nucleotides from the 5' end of RNA to locate the cleavage site.**

**a**, Processing of pre-let-7a-1 (Let-7) with extra 3' uridine tails using immunopurified human Dicer. Numbers on top of RNA substrates indicate the length of 3' overhang. Purple and green arrowheads indicate the 5' counting and 3' counting cleavage products, respectively (this standard is used throughout the figures). Asterisks mark radio-labelled phosphates, throughout the figures. nt, nucleotide. **b**, Dicer processing of dsRNAs (ds-35) with extra 3' uridine tails. Grey arrowhead indicates the cleavage product from the opposite end. **c**, The 5' end of ds-35 was extended by 1 nucleotide (ds-35+5'C) and incubated with Dicer. **d**, Dicer processing of the ds-35 substrates with either a 5'-terminal phosphate or a hydroxyl group (cyan circle). Although the juxtaposed lanes are not contiguous, they are all from a single gel; this is true for all images with dashed lines throughout the figures.

substrates most strongly by using both 3'-dependent and 3'-independent mechanisms. When the terminal structure deviates from the optimal 2-nucleotide 3' overhang, either one of the two mechanisms seems to be used by Dicer, yielding the mixture of two distinct product populations.

We next questioned what determines the 3'-independent selection of the cleavage site. Because the cleavage site is always 22 nucleotides away from the 5' end, we presumed that the 5' end may have a role. To test this idea, dsRNA substrates were extended by adding one cytidine at the 5' end (Fig. 1c). Dicer still yielded products of ~22 nucleotides, indicating that the cleavage site shifted by 1 nucleotide when the 5' end was extended. Hence, Dicer measures a set distance from the 5' end. We refer to this as the '5' counting model'.

Because endogenous substrates of Dicer carry a 5'-terminal phosphate group, we tested whether the 5' phosphate has a role in the recognition of the 5' end. Dicer processing was performed with two sets of dsRNA substrates that carry either a 5'-terminal phosphate or a hydroxyl group (Fig. 1d). Both sets were labelled at the 3' end of the opposite terminus to detect the cleavage products. The cleavage patterns of the two sets differed markedly. The phosphorylated dsRNAs followed the 5' counting rule whereas the dsRNAs lacking the 5' phosphate mainly obeyed the 3' counting rule (Fig. 1d). We also noticed that the length of the products became more variable in the absence of the 5' phosphate. Thus, Dicer interaction with the 5'-terminal phosphate helps precisely locate the enzyme on the substrate.

To investigate whether the 5' counting rule can be generalized to other pre-miRNAs, we examined additional pre-miRNAs with 1–3 extra uridine residues at the 3' end (Supplementary Fig. 2). Pre-miR-143, pre-miR-148b, pre-miR-27b and pre-miR-151 largely followed the 5' counting rule, whereas pre-miR-200c showed a mixed pattern. Pre-miR-24-2 and pre-miR-142 complied mainly with the 3' counting rule. Hence, although the 5' counting applies to most pre-miRNAs tested, the relative contribution of the 5' and 3' ends seems to vary among pre-miRNAs. We noticed that pre-miRNAs following the 3' counting rule are relatively stable at the stem termini, whereas pre-miRNAs following the 5' counting rule have less stable structures at the terminal base pair (mismatch, G–U, or A–U pair) (Supplementary Fig. 2a). Thus, Dicer may require a flexible (thermodynamically unstable) 5' terminus to efficiently recognize the 5' end. To test this notion further, we changed  $Mg^{2+}$  concentrations in our processing assays because the  $Mg^{2+}$  ion is known to stabilize the dsRNA structure<sup>23</sup>.  $Mg^{2+}$  ions indeed had a significant influence on Dicer processing of pre-miR-24-2: the 3' counting rule prevails at 4 mM whereas the 5' counting rule predominates at 0.5 mM (Supplementary Fig. 3a). A similar observation was made when a duplex RNA with a 3-nucleotide overhang was used (Supplementary Fig. 3b, c). It is likely that at a low  $Mg^{2+}$  concentration, the terminal stem region tends to unwind, thereby facilitating the 5'-end recognition by Dicer. Given that the physiological concentration of free  $Mg^{2+}$  ions is estimated to be 0.5–1 mM<sup>24</sup>, the 5' counting rule may apply to most 3'-modified

pre-miRNAs *in vivo*, although we do not exclude the possibility that some pre-miRNAs with stable termini may follow the 3' counting rule.

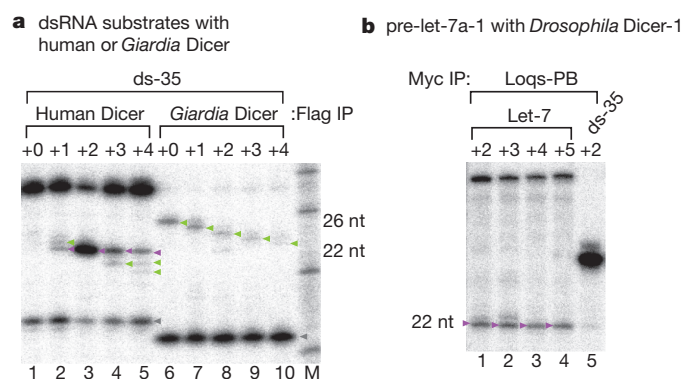
### Conservation of the 5' counting rule

As the 3' counting model was proposed mainly from the work on *Giardia* Dicer<sup>15,16</sup>, we examined *Giardia* Dicer using our substrates (Fig. 2a, lanes 6–10). Unlike human Dicer, *Giardia* Dicer cleaved dsRNA substrates by strictly measuring from the 3' end and yielded slightly larger products (24–26 nucleotides), as previously observed<sup>16</sup>. We also noticed that *Giardia* Dicer cleaves the blunt-ended substrate most efficiently, whereas human Dicer shows a strong preference for the 2-nucleotide 3' overhang terminal structure. Thus, *Giardia* Dicer differs significantly from human Dicer in substrate recognition.

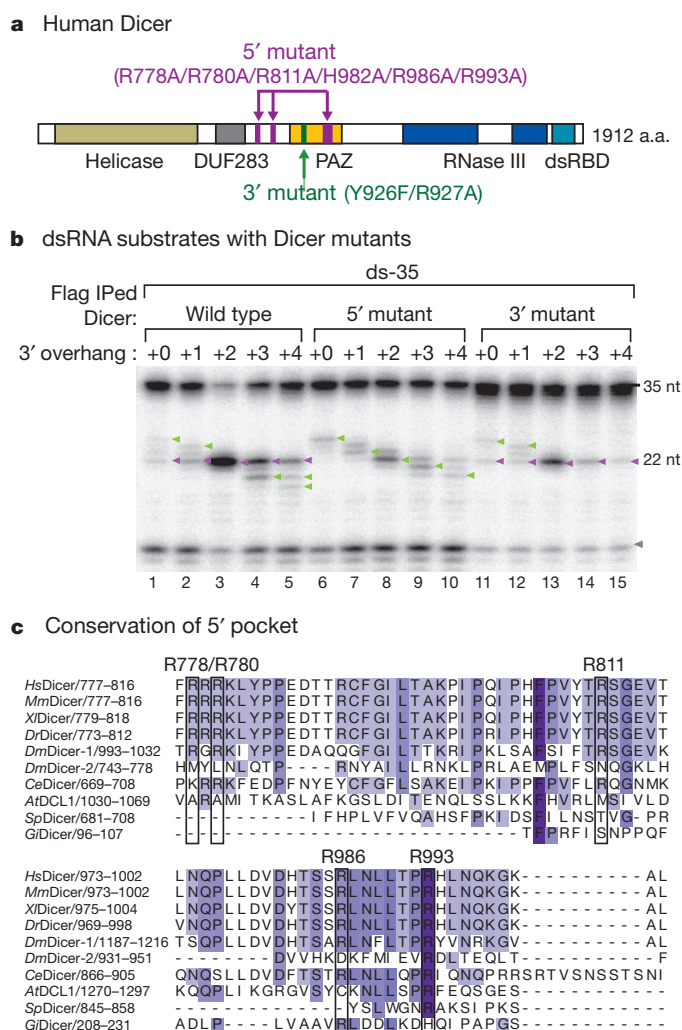
We next investigated the processing pattern of *Drosophila* Dicer-1. Dicer-1 acts in complex with the co-factor Loquacious-PB (Loqs-PB) for the processing of pre-miRNAs<sup>25–28</sup>, while another Dicer (Dicer-2) and its cofactor R2D2 are responsible for siRNA generation<sup>29</sup>. When pre-let-7a-1 variants were incubated with the Dicer-1–Loqs-PB complex, all variants were cleaved into 22-nucleotide products, without any detectable products following the 3' counting (Fig. 2b). Taken together, the 5'-end recognition mechanism may be conserved in metazoans but not in organisms such as *Giardia*, which represents one of the earliest surviving branches of the eukaryotic phylogenetic tree.

### Identification of the 5'-recognition pocket

To identify the motif that binds to the 5' end, we selected putative RNA interacting residues (basic, polar) located around the PAZ domain and mutated them to alanines (Fig. 3a). The residues were selected on the basis of three criteria. First, we predicted the three-dimensional structure of the region encompassing the PAZ domain by applying I-TASSER simulation<sup>30</sup> (Supplementary Fig. 4). Assuming that the 5'-end-binding residues are located ~20 Å away from the conserved 3'-end-binding pocket (3' pocket), which is the expected distance between the 5' and 3' ends of a 2-nucleotide 3' overhang structure<sup>14</sup>, we selected putative RNA interacting residues (R811, R986 and R993). Second, we solved the crystal structure of a human Dicer fragment spanning the 'platform–PAZ–connector–helix' domains (Supplementary Fig. 5 and manuscript in preparation). Interestingly, we found an inorganic phosphate that is coordinated to the side chains of R778, R780, R811 and H982, suggestive of a potential phosphate-binding pocket. The bound phosphate is located ~20 Å away from the 3' pocket in the PAZ domain. R986 and R993 are in a disordered part of the structure (Supplementary Fig. 5a) but they are predicted to be in



**Figure 2 | The 5' counting rule is conserved in *Drosophila* Dicer-1 but not in *Giardia* Dicer.** **a**, Flag-tagged human Dicer and *Giardia* Dicer were immunopurified (IP) and incubated with ds-35 substrates. **b**, Immunopurified *Drosophila* Dicer-1–Loqs-PB complexes were incubated with Let-7 substrates. The ds-35 (+2) substrate was used as a negative control.



**Figure 3 | Residues required for the recognition of dsRNA terminus.**

**a**, Domain organization of human Dicer. The mutated sites in the 5' and 3' pockets are shown as purple and green bars, respectively, and the mutations are listed. a.a., amino acids. **b**, Immunopurified wild-type and mutant Dicer proteins were incubated with ds-35 substrates. **c**, Amino acid sequences of the Dicer proteins from various species (*At*, *Arabidopsis thaliana*; *Ce*, *Caenorhabditis elegans*; *Dm*, *Drosophila melanogaster*; *Gi*, *Giardia intestinalis*; *Hs*, *Homo sapiens*; *Mm*, *Mus musculus*; *Sp*, *Schizosaccharomyces pombe*; *Xl*, *Xenopus laevis*) are aligned using ClustalX program and the region spanning the 5' pocket is presented. The 5'-interacting residues are indicated with boxes.

the vicinity of R778. Finally, in deciding on the residues for the mutagenesis study, we took into account phylogenetic conservation.

The mutants at R778/R780, R811 and R986/R993 produced a significantly smaller amount of 5' counting products, indicating that these mutants are defective in 5'-end recognition (Supplementary Fig. 6a). When we combined these mutations to generate a '5' mutant' (R778A/R780A/R811A/H982A/R986A/R993A), the cleavage pattern clearly shifted to the 3' counting one (Fig. 3b, lanes 6–10). The change in the cleavage pattern is highly specific to the identified residues; the point mutations at S984, H994 and W1014, which are located closely to the 5'-interacting residues, did not affect the cleavage pattern (Supplementary Fig. 6b). Thus, our results indicate that a basic motif composed of R778, R780, R811, R986 and R993 (5' pocket) is required for 5'-end recognition. These amino acids are conserved in *Drosophila* Dicer-1 but not in *Giardia* Dicer (Fig. 3c).

As a control, we introduced mutations at Y926 and R927, which are conserved and located in the 3' pocket of the PAZ domain (Fig. 3a). This mutant (3' mutant) lost most of the 3' counting products, indicating that the 3' counting mechanism is disrupted in this mutant



(Fig. 3b, lanes 11–15). This result is consistent with previous findings that 3' counting is dependent on the interaction between the 3' end and the PAZ domain<sup>14–16</sup>.

Overall processing efficiency was reduced in the 5' mutant as well as in the 3' mutant (Fig. 3b), consistent with the notion that human Dicer utilizes both ends for substrate binding. The substrate with a 2-nucleotide overhang was cleaved more heterogeneously by the 5'-mutant Dicer (21–23 nucleotides) compared to wild-type Dicer (22 nucleotides) (Fig. 3b, compare lanes 3 and 8), indicating that the 5' mutant lacks precision in processing. Altogether, our *in vitro* data indicate that 5'-end recognition is important not only for 3'-modified pre-miRNAs but also for canonical substrates such as unmodified pre-miRNAs.

### The 5' pocket is required for miRNA biogenesis

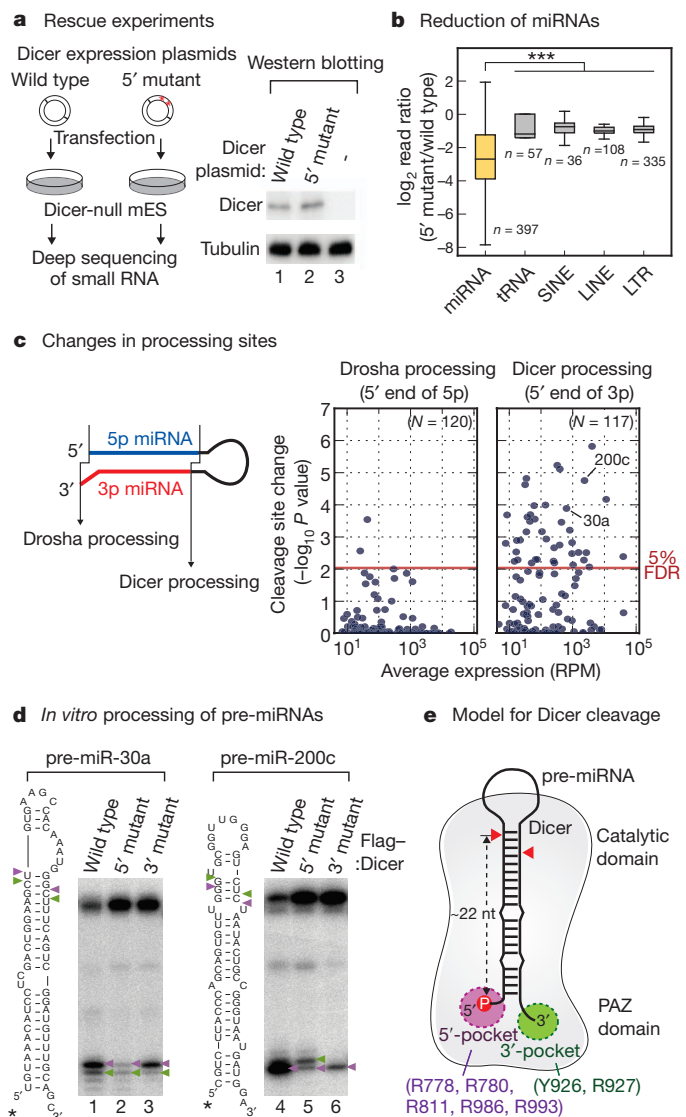
To evaluate the biological relevance of our findings, we introduced Dicer expression plasmids (wild type and 5' mutant) into Dicer-null embryonic stem (ES) cells<sup>31</sup>. The small RNA populations from two biological replicates were sequenced (Fig. 4a and Supplementary Table 1). Wild-type Dicer successfully replenished the miRNA pool whereas the 5' mutant showed significant defects. The overall miRNA abundance decreased in the 5'-mutant-expressing cells (Fig. 4b, Supplementary Fig. 7 and Supplementary Table 2), although some miRNA isoforms increased owing to cleavage site alterations (see later). Other RNA species such as transfer RNAs were unaffected, indicating that the differences are specific to miRNAs. The marked reduction of miRNA abundance was further confirmed by northern blotting (Supplementary Fig. 7d).

We next examined the impact of the 5'-pocket mutation on processing site selection. As the 3' ends of small RNAs are known to be frequently modified after Dicer processing<sup>17,20</sup>, we used the 5' ends of miRNAs (or miRNAs\*) to infer cleavage sites. Drosha creates the 5' end of 5'-strand miRNAs (5p miRNAs) whereas Dicer makes the 5' end of 3'-strand miRNAs (3p miRNAs) (Fig. 4c, left). When the wild-type and 5'-mutant libraries were compared, ~35% of miRNAs showed significant changes in Dicer cleavage sites (41 out of 117; below 5% false discovery rate) (Fig. 4c, right panel, Supplementary Figs 8 and 9 and Supplementary Table 2). In contrast, Drosha processing sites remained largely unchanged (Fig. 4c, left panel, and Supplementary Fig. 9), indicating that the differences in the small RNA population are due to the mutation in Dicer. The changes in Dicer cleavage sites often led to seed alterations and/or strand switches (Supplementary Table 3). The deep sequencing results are highly consistent with those from *in vitro* assays.

As further confirmation, we carried out *in vitro* processing of pre-miR-30a and pre-miR-200c, which showed significant changes in the 5'-mutant-expressing cells (Fig. 4c and Supplementary Fig. 10). The 5'-mutant Dicer was markedly impaired in both efficiency and accuracy of pre-miRNA processing *in vitro* (Fig. 4d). The 3'-pocket mutation reduced processing activity without significantly altering cleavage site selectivity (Fig. 4d). Taken together, the 5' pocket is critical for efficient and precise generation of miRNA.

### Discussion

This study provides new insight into the mechanism of Dicer processing (see Fig. 4e for a model). The basic 5' pocket identified in our study is positioned in close proximity to the 3' pocket on the same face of the Dicer protein. The 5' and 3' pockets of Dicer are positioned for the simultaneous accommodation of the 5' and 3' ends, respectively, of the substrate with a 2-nucleotide 3' overhang. The 5' pocket anchoring the 5' phosphate is particularly important for securing Dicer in a fixed position, which enables Dicer to generate uniform products. Ongoing structural studies on RNA complexes of PAZ-containing fragments of metazoan Dicer will elucidate further the molecular basis of Dicer processing.



**Figure 4 | The 5' pocket is critical for efficient and accurate cleavage *in vivo*.** **a**, Left, scheme for the Dicer rescue experiment using Dicer-null mouse ES cells. Right, western blotting using anti-Dicer antibody shows the comparable expression of wild-type and mutant Dicer. Tubulin was used as a loading control. **b**, Transcripts mapped to miRNA loci specifically decreased in the mutant library ( $P = 8.13 \times 10^{-59}$ , Mann–Whitney *U*-test, first replicate). Box represents the first and third quartiles and the internal bar indicates the median. Whiskers denote the lowest and highest values within  $1.5 \times$  interquartile range of the first and third quartiles, respectively. *n* represents the number of transcripts. **c**, Left, Drosha and Dicer cleavage sites were inferred from the 5' end of 5p and 3p miRNAs, respectively. Right, the dissimilarity of cleavage pattern was quantified by Kullback–Leibler divergence (KLD), and statistical significance was measured using two-sample *t*-test. Red line corresponds to 5% false discovery rate (FDR). RPM, reads per million. **d**, Pre-miR-30a and pre-miR-200c were incubated with the same amount of wild-type or mutant Dicer protein. **e**, Double anchor model for Dicer processing.

It is interesting to contemplate the evolutionary implications because the 5'-pocket motif is highly conserved among most miRNA-producing Dicer homologues. This motif is missing in Dicer from lower eukaryotes such as *Giardia* and fungi (Fig. 3c), which lack the miRNA pathway. Plant DCL1, the miRNA-producing enzyme, is only partially conserved in this region. Therefore, on the basis of the amino acid sequences, it is difficult to infer the existence of an orthologous motif. It is notable that, in *Drosophila*, the 5'-pocket motif seems to be conserved only in Dicer-1 (miRNA-producing enzyme) but not in Dicer-2 (siRNA-generating enzyme). The fact that

the 5' pocket is conserved only in Dicers with pre-miRNA processing activity implicates the 5' counting mechanism as an important factor in miRNA maturation.

In the cell, 3' ends of pre-miRNAs can be modified by exonucleases or nucleotidyl transferases<sup>17–20,32</sup>. It is yet unclear to what extent the 5'-pocket motif contributes to the processing of such modified pre-miRNAs *in vivo*, as we would need to determine exactly what fractions of pre-miRNAs undergo 3'-end modifications *in vivo*. Nonetheless, because the 5' end of pre-miRNAs is generally more homogenous than the 3' end<sup>33</sup>, it is tempting to speculate that the 5' pocket has evolved to utilize the 5' end, which is more reliable than the 3' end. Using the 5' end as a major reference point for positioning may ensure accurate processing of pre-miRNAs. In the case of the siRNA pathway, precise processing is not critical because, unlike miRNAs, any cleavage frame would result in functional siRNAs.

RNA interference (RNAi) in mammalian systems is commonly induced by expressing small hairpin RNAs (shRNAs) from an RNA polymerase II or III promoter, but the technology often suffers from inefficient and inaccurate Dicer processing<sup>34–36</sup>. On the basis of our findings, a hairpin with a 5'-terminal phosphate and a 2-nucleotide 3' overhang should fit most optimally into the 5' and 3' pockets of Dicer. Also, it would be interesting to test whether a 5' triphosphate could be efficiently accommodated into the 5' pocket, as shRNAs driven by RNA polymerase III promoters bear a 5' triphosphate. Understanding how human Dicer generates miRNAs will enable us to improve further the efficacy and safety of RNAi technology.

## METHODS SUMMARY

*In vitro* processing was performed by incubating end-labelled RNA with immunopurified Dicer proteins. Flag-tagged human or *Giardia* Dicer was expressed in HEK293T cells. *Drosophila* Dicer-1 complex was purified by immunoprecipitation of Myc-tagged Loqs-PB from S2 cells. RNA substrates were either chemically synthesized or generated by ligating two synthetic single-stranded RNAs. For *in vivo* experiments, the 5'-mutant Dicer construct was transfected into Dicer-null mouse ES cells. The small RNA population from the mouse ES cells was analysed by Illumina deep sequencing.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 13 September 2010; accepted 16 May 2011.

- Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. *Nature Rev. Mol. Cell Biol.* **10**, 126–139 (2009).
- Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
- Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659 (2001).
- Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
- Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34 (2001).
- Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* **293**, 834–838 (2001).
- Knight, S. W. & Bass, B. L. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* **293**, 2269–2271 (2001).
- Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150 (2001).
- Tabara, H., Yigit, E., Siomi, H. & Mello, C. C. The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DEXH-box helicase to direct RNAi in *C. elegans*. *Cell* **109**, 861–871 (2002).
- Mourelatos, Z. *et al.* miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* **16**, 720–728 (2002).
- Han, J. *et al.* Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**, 887–901 (2006).
- Vermeulen, A. *et al.* The contributions of dsRNA structure to Dicer specificity and efficiency. *RNA* **11**, 674–682 (2005).

- Zhang, H., Kolb, F. A., Brondani, V., Billy, E. & Filipowicz, W. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.* **21**, 5875–5885 (2002).
- Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E. & Filipowicz, W. Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**, 57–68 (2004).
- MacRae, I. J., Zhou, K. & Doudna, J. A. Structural determinants of RNA recognition and cleavage by Dicer. *Nature Struct. Mol. Biol.* **14**, 934–940 (2007).
- MacRae, I. J. *et al.* Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195–198 (2006).
- Burroughs, A. M. *et al.* A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. *Genome Res.* **20**, 1398–1410 (2010).
- Wu, H., Ye, C., Ramirez, D. & Manjunath, N. Alternative processing of primary microRNA transcripts by Drosha generates 5' end variation of mature microRNA. *PLoS ONE* **4**, e7566 (2009).
- Heo, I. *et al.* Lin28 mediates the terminal uridylation of *let-7* precursor microRNA. *Mol. Cell* **32**, 276–284 (2008).
- Chiang, H. R. *et al.* Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev.* **24**, 992–1009 (2010).
- Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
- Tomari, Y., Matranga, C., Haley, B., Martinez, N. & Zamore, P. D. A protein sensor for siRNA asymmetry. *Science* **306**, 1377–1380 (2004).
- Serra, M. J. *et al.* Effects of magnesium ions on the stabilization of RNA oligomers of defined structures. *RNA* **8**, 307–323 (2002).
- Gunther, T. Concentration, compartmentation and metabolic function of intracellular free  $Mg^{2+}$ . *Magnes. Res.* **19**, 225–236 (2006).
- Jiang, F. *et al.* Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev.* **19**, 1674–1679 (2005).
- Saito, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in *Drosophila* cells. *PLoS Biol.* **3**, e235 (2005).
- Miyoshi, K., Miyoshi, T., Hartig, J. V., Siomi, H. & Siomi, M. C. Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in *Drosophila*. *RNA* **16**, 506–515 (2010).
- Förstemann, K. *et al.* Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* **3**, e236 (2005).
- Lee, Y. S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).
- Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* **5**, 725–738 (2010).
- Murchison, E. P., Partridge, J. F., Tam, O. H., Cheloufi, S. & Hannon, G. J. Characterization of Dicer-deficient murine embryonic stem cells. *Proc. Natl Acad. Sci. USA* **102**, 12135–12140 (2005).
- Heo, I. *et al.* TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* **138**, 696–708 (2009).
- Seitz, H., Ghildiyal, M. & Zamore, P. D. Argonaute loading improves the 5' precision of both microRNAs and their miRNA\* strands in flies. *Curr. Biol.* **18**, 147–151 (2008).
- Chang, K., Elledge, S. J. & Hannon, G. J. Lessons from Nature: microRNA-based shRNA libraries. *Nature Methods* **3**, 707–714 (2006).
- Silva, J., Chang, K., Hannon, G. J. & Rivas, F. V. RNA-interference-based functional genomics in mammalian cells: reverse genetics coming of age. *Oncogene* **23**, 8401–8409 (2004).
- Kim, D. H. & Rossi, J. J. Strategies for silencing human disease using RNA interference. *Nature Rev. Genet.* **8**, 173–184 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to G. Hannon for Dicer-null mouse ES cells; J. Doudna for *Giardia* Dicer cDNA; and M. Siomi for *Drosophila* Dicer-1, Dicer-2, Loqs-PB and R2D2 constructs. We also thank the members of the V.N.K. laboratory, particularly C. Joo, M.-J. Yoon, K.-H. Yeom and A. Cho for discussions and technical help. The V.N.K. laboratory was supported by the Creative Research Initiatives Program (2010000021) and National Honor Scientist Program (20100020415) through the National Research Foundation and the BK21 Fellowships (J.-E.P. and H.C.) from the Ministry of Education, Science and Technology. Research in the D.J.P. laboratory was supported by the National Institutes of Health.

**Author Contributions** J.-E.P., I.H. and D.J. performed biochemical and cell biological experiments. H.C. carried out bioinformatic analyses. Y.T., D.K.S. and D.J.P. performed structural studies. J.-E.P., I.H. and V.N.K. designed the study and wrote the paper.

**Author Information** Small RNA sequencing data were deposited in the Gene Expression Omnibus ([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)) under accession number GSE27903. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to V.N.K. ([narrykim@snu.ac.kr](mailto:narrykim@snu.ac.kr)).



## METHODS

**Cell culture and transfection.** HEK293T cells were grown in DMEM (Welgene) supplemented with 10% fetal bovine serum (Welgene). S2 cells were grown in HyClone SFX-Insect (Thermo Scientific) supplemented with 10% fetal bovine serum (Welgene). Dicer knockout mouse ES cells (a gift from G. J. Hannon) were grown on mouse CF-1 feeder cells or gelatin-coated dishes in knockout DMEM (Gibco, Invitrogen) supplemented with 15% fetal bovine serum (Gibco, Invitrogen), nonessential amino acids (Gibco, Invitrogen), 2 mM L-glutamine (Sigma), 0.1 mM 2-mercaptoethanol (Sigma) and 1,000 units ml<sup>-1</sup> leukaemia inhibitory factor (Chemicon).

For HEK293T cells, transfection was carried out using the calcium-phosphate method. S2 cells were transfected using the DDAB method as previously described<sup>37</sup>.

**Cloning and mutagenesis.** *Giardia* Dicer cDNA was amplified from *Giardia* Dicer-pFastBac HTa plasmid (a gift from J. A. Doudna) by PCR using the following primers: 5'-GGATCCATGCATGCTTTGGGACACTG-3'; and 5'-GATATCGAGACTGCAGGCTCTAGATTTCG-3'. PCR products were cloned into pGEM-T easy vector (Promega) and subsequently cloned into Flag-pcDNA3 vector (Invitrogen) at the BamHI and EcoRV site.

To introduce mutations into Dicer, QuickChange Site-Directed Mutagenesis Kit (Stratagene) was used. Mutated plasmids were confirmed by sequencing and subcloned into unmodified Flag-Dicer-pcDNA3.1 vector. The primer sequences used for the mutagenesis are provided in Supplementary Table 4.

**Immunoprecipitation and *in vitro* Dicer processing.** For immunoprecipitation of Flag-Dicer, HEK293T cells were grown on 10-cm or 15-cm dishes and harvested at 48 h after Flag-Dicer-pcDNA3.1 expression plasmid transfection. The cells were incubated with lysis buffer (500 mM NaCl, 1 mM EDTA, 20 mM Tris (pH 8.0), 1% Triton X-100) for 20 min on ice followed by sonication and centrifugation twice at 16,000g for 10 min at 4 °C. The supernatant was incubated with 10 µl of anti-Flag antibody conjugated to agarose beads (anti-Flag M2 affinity gel, Sigma) with constant rotation for 1 h at 4 °C. The beads were washed three times with lysis buffer and then four times with buffer D (200 mM KCl, 20 mM Tris (pH 8.0), 0.2 mM EDTA). The reactions were performed in a total volume of 30 µl in 2 mM MgCl<sub>2</sub>, 1 mM DTT, 1 unit µl<sup>-1</sup> ribonuclease inhibitor (Takara), 5'-end-labelled pre-miRNA of 1 × 10<sup>4</sup> to 1 × 10<sup>5</sup> c.p.m. and 15 µl of the immunopurified proteins in buffer D. The reaction mixture was incubated at 37 °C for 60–90 min. RNA was purified from the reaction mixture by phenol extraction and separated on 15% urea polyacrylamide gel. Along with Decade marker (Ambion), synthetic hsa-let-7a RNA (22 nucleotides) was 5'-end labelled and used as a size marker, because the 20-nucleotide RNA in Decade marker is often degraded to 18–19 nucleotides as we previously reported<sup>11</sup>.

For preparation of *Drosophila* Dicer-1, S2 cells confluent in a 10-cm dish were transfected with Myc-Loquacious-PB-pRmHa3 expression plasmid (a gift from M. C. Siomi). We used the Loquacious-PB immunoprecipitates instead of the Dicer-1 immunoprecipitates in this experiment because the ectopic expression level of Dicer-1 was too low. After 1 day, 1 mM CuSO<sub>4</sub> was added to the medium and cells were collected 2 days after CuSO<sub>4</sub> treatment. The cells were incubated with lysis buffer for 30 min on ice, followed by sonication and centrifugation at 16,000g for 10 min at 4 °C. The supernatants were pre-cleared by incubation with 10 µl Protein A-Sepharose bead (GE Healthcare) for 2 h. Then, pre-cleared extract was incubated with 20 µl Protein A-Sepharose bead bound to anti-Myc antibody, 9E10, for 2 h at 4 °C. The beads were washed three times with lysis buffer and then four times with buffer D and used for *in vitro* Dicer processing.

**Preparation of substrates.** Pre-let-7a-1, pre-miR-16-1, pre-miR-24-2 (mouse), pre-miR-142, pre-miR-143, pre-miR-200c and pre-miR-30a were synthesized by ST Pharm. The sequences are presented in the figures. The pre-miRNA substrates with different 3'-overhang lengths and pre-miR148b, pre-miR-27b and pre-miR-151 were generated by ligating two synthetic ssRNAs as described previously<sup>32</sup>. The sequences of RNA used for ligation are listed in Supplementary Table 5. The RNAs were labelled at the 5' end with T4 polynucleotide kinase (T4 PNK, Takara) and [ $\gamma$ -<sup>32</sup>P] ATP. Sequences of all endogenous pre-miRNAs used in our analysis are listed in Supplementary Table 6.

For preparation of dsRNA substrates, a synthetic ssRNA was labelled at the 5' end with [ $\gamma$ -<sup>32</sup>P] ATP and T4 PNK. After phenol extraction, the labelled RNA was annealed to the complementary RNA by heating at 90 °C for 2 min and incubating at 30 °C for 2 h. In Fig. 1d, one strand of RNA was ligated to [ $\alpha$ -<sup>32</sup>P] pCp and treated with calf intestinal alkaline phosphatase (Takara) to generate the labelled 3' end with a hydroxyl group. To attach a phosphate group at the 5' end, 3'-end-labelled RNAs were incubated with cold ATP and T4 PNK (Takara). Phenol extraction of RNA was performed after each reaction. Then, the labelled RNA was annealed to the RNA as described earlier.

***In vitro* addition of uridine residues to pre-miRNAs.** A terminal nucleotidyl transferase, TUT4, is able to add 1–3 nucleotides of uridine residues at the 3' end of

pre-miRNA in the absence of Lin28 protein *in vitro* (I. Heo *et al.*, unpublished data). For this reaction, Flag-TUT4 expression plasmid was transfected in HEK293T cells using the calcium-phosphate method. After 48 h, total cell extract was prepared in buffer D by sonication and centrifugation at 16,000g for 10 min at 4 °C. Thirty microlitres of reaction mixture contains 15 µl total cell extract (10 µg), 3.2 mM MgCl<sub>2</sub>, 1 mM DTT, 0.25 mM UTP, 1 unit µl<sup>-1</sup> ribonuclease inhibitor (Ambion), and 5'-end-labelled pre-miRNA of 1 × 10<sup>4</sup> to 1 × 10<sup>5</sup> c.p.m. The reaction mixture was incubated at 37 °C for 15 min. After phenol extraction, the uridylylated pre-miRNAs were gel purified and used for *in vitro* Dicer processing.

**Three-dimensional structure prediction of Dicer fragment.** The three-dimensional structure of Dicer fragment containing the PAZ domain was predicted by I-TASSER simulation<sup>30</sup> (<http://zhanglab.ccmb.med.umich.edu/I-TASSER>) with amino acid sequences 751–1070. Crystal structure of PAZ domains from Dicer (PDB accession code 2FFL) and Argonautes (PDB codes 1U04, 3DLB, 1R4K), together with RumA, a 23S ribosomal RNA methyltransferase (PDB code 1UWV), were used as templates for the comparative modelling. Among the five models predicted from the server, the one with a high C score (−2.36) and an organized structure was chosen.

**Structure-based identification of the 5' pocket in human Dicer.** Diffraction quality crystals were grown for the complex of Dicer 'platform-PAZ-connector-helix' cassette (residues 755–1055) and a self-complementary AGCGAAUUCGCUU duplex (underlined segment forms duplex) in phosphate-containing solution. The crystals of the complex belonged to space group I222, diffracted to 2.6 Å, and the structure of the complex was refined to  $R_{\text{work}} = 19.7$  and  $R_{\text{free}} = 23.7$ . In this structure, inorganic phosphate, which is anchored by basic residues (Arg 778, Arg 780, Arg 811 and His 982), reveals the potential 5'-phosphate-binding pocket (Supplementary Fig. 5).

**Dicer rescue experiments.** For transfection, Dicer knockout mouse ES cells were separated from feeder cells and 1,500,000 cells were seeded on gelatin-coated 6-well plates one day before transfection. Ten micrograms of plasmids (wild-type Dicer-pCK or 5'-mutant Dicer-pCK) were added to each well along with 10 µl of Lipofectamine 2000, according to the manufacturer's protocol (Invitrogen). Protein and RNA was extracted at 48 h after transfection. To determine the protein levels, western blotting was performed using anti-Dicer and anti-tubulin (Abcam) antibodies. Expression of RNA was confirmed by northern blotting using the following probes: mmu-miR-293 (5'-ACACTACAACTCTGCGGCACT-3'); mmu-miR-101a (5'-TTCAGTTATCACAGTACTGTA-3'); mmu-miR-16 (5'-CGCCAATATTTACGTGCTGCTA-3'); and tRNA-Lys-AAG (5'-GAGATTAAGAGTCTCATGCTC-3').

To prepare small RNA cDNA libraries, RNA was extracted using TRIzol reagent (Invitrogen) or mirVana miRNA isolation kit (Ambion) and separated on 15% urea-PAGE. RNA of 17–26 nucleotides in length was gel purified and ligated to the 3' adaptor using truncated T4 RNA Ligase2 (NEB) in ATP-free conditions. Subsequently, the ligation product was gel purified and ligated to the 5' adaptor using T4 RNA Ligase1 (NEB). The final ligation product was gel purified and used for reverse transcription using SuperScript II (Invitrogen). The cDNA was PCR-amplified with Phusion DNA polymerase (NEB). The resulting libraries were sequenced using Illumina Genome Analyser II.

**Sequence analysis.** The essential workflow for early sequence analysis was performed as previously described<sup>38</sup> with few modifications. After removing sequence reads including very low quality bases (<10 in phred quality), the 3' adaptor sequence was trimmed from the reads using a 5'-free variant of the Smith-Waterman algorithm (scoring parameters: 2 for match, −3 for mismatch, −3 for linear gap). Then, we dropped short (≤17 nucleotides) or repetitive sequences (0.7 and 1.5 for mono- or dinucleotide entropy of each sequence). The filtered sequences were aligned to Illumina adaptor and primer sequences using the BWA short-read aligner<sup>39</sup> with 4 of allowed maximum edit distance, then matched reads were removed from further analysis. In the same way, the remaining sequences were aligned to the mouse genome mm9 assembly, which is downloaded from the University of California at Santa Cruz (UCSC; <http://genome.cse.ucsc.edu/>). Annotations for aligned regions were retrieved using in-house software from RefSeq, RepeatMasker and miRBase (downloaded from UCSC or miRBase on April 8, 2011). Software used in data processing and analysis can be downloaded from <http://www.narrykim.org/s/park-dicer-2011/>.

**Analysis of cleavage site change.** We first selected miRBase stem-loops that are relatively unaffected by reads aligned to multiple miRNA loci to avoid artefacts from over- or underestimated read counts. Stem-loops with more than 90% reads aligned to a single stem-loop in every single sequencing lane were chosen for the later steps. Kullback-Leibler divergence (KLD) was used to quantify cleavage site change (difference of 5'-end position frequency) between two sequencing samples. To measure statistical significance of cleavage site change, Student's *t*-test was performed for KLDs between wild-type and 5'-mutant Dicer rescued samples, and KLDs between wild-type Dicer rescued samples and J1<sup>40</sup>, mouse embryo at

- 7.5 day<sup>20</sup> or R1. Multiple testing correction was applied using the Benjamini–Hochberg method<sup>41</sup>.
37. Han, K. An efficient DDAB-mediated transfection of *Drosophila* S2 cells. *Nucleic Acids Res.* **24**, 4362–4363 (1996).
  38. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
  39. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
  40. Babiarz, J. E., Ruby, J. G., Wang, Y., Bartel, D. P. & Blelloch, R. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev.* **22**, 2773–2785 (2008).
  41. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J. R. Stat. Soc., B* **57**, 289–300 (1995).



# A low mass for Mars from Jupiter's early gas-driven migration

Kevin J. Walsh<sup>1,2</sup>, Alessandro Morbidelli<sup>1</sup>, Sean N. Raymond<sup>3,4</sup>, David P. O'Brien<sup>5</sup> & Avi M. Mandell<sup>6</sup>

Jupiter and Saturn formed in a few million years (ref. 1) from a gas-dominated protoplanetary disk, and were susceptible to gas-driven migration of their orbits on timescales of only  $\sim 100,000$  years (ref. 2). Hydrodynamic simulations show that these giant planets can undergo a two-stage, inward-then-outward, migration<sup>3–5</sup>. The terrestrial planets finished accreting much later<sup>6</sup>, and their characteristics, including Mars' small mass, are best reproduced by starting from a planetesimal disk with an outer edge at about one astronomical unit from the Sun<sup>7,8</sup> (1 AU is the Earth–Sun distance). Here we report simulations of the early Solar System that show how the inward migration of Jupiter to 1.5 AU, and its subsequent outward migration, lead to a planetesimal disk truncated at 1 AU; the terrestrial planets then form from this disk over the next 30–50 million years, with an Earth/Mars mass ratio consistent with observations. Scattering by Jupiter initially empties but then repopulates the asteroid belt, with inner-belt bodies originating between 1 and 3 AU and outer-belt bodies originating between and beyond the giant planets. This explains the significant compositional differences across the asteroid belt. The key aspect missing from previous models of terrestrial planet formation is the substantial radial migration of the giant planets, which suggests that their behaviour is more similar to that inferred for extrasolar planets than previously thought.

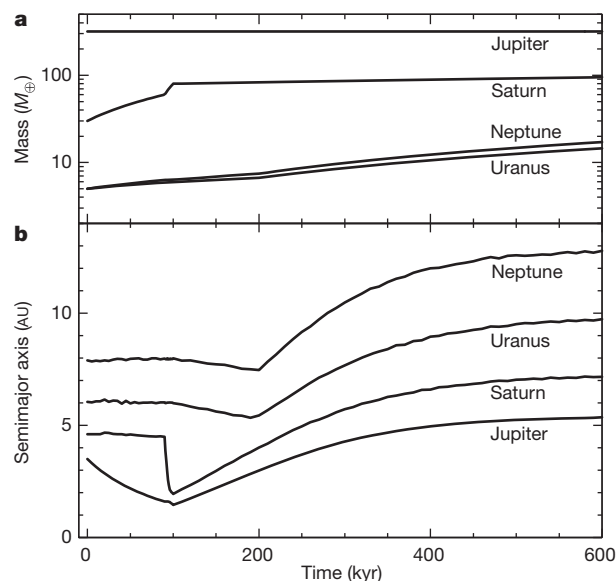
Hydrodynamic simulations show that isolated giant planets embedded in gaseous protoplanetary disks carve annular gaps and migrate inward<sup>9</sup>. Saturn migrates faster than Jupiter; if Saturn is caught in the 2:3 mean motion resonance with Jupiter (conditions for this to happen are given in Supplementary Information section 3), where their orbital period ratio is 3/2, generally the two planets start to migrate outward until the disappearance of the disk<sup>3–5,10</sup>. Jupiter could have migrated inward only before Saturn approached its final mass and was captured in resonance. The extents of the inward and outward migrations are unknown a priori owing to uncertainties in disk properties and in relative timescales for the growth of Jupiter and Saturn. Thus we search for constraints on where Jupiter's migration may have reversed (or 'tacked', using a sailing analogy).

The terrestrial planets are best reproduced when the disk of planetesimals from which they form is truncated, with an outer edge at 1 AU (refs 7, 8). These conditions are created naturally if Jupiter tacked at  $\sim 1.5$  AU. However, before concluding that Jupiter tacked at this distance, a major question needs to be addressed: can the asteroid belt, between 2 and 3.2 AU, survive the passage of Jupiter?

Volatile-poor asteroids (mostly S types) are predominant in the inner asteroid belt, while volatile-rich asteroids (mostly C types) are predominant in the outer belt. These two main classes of asteroids have partially overlapping semimajor axis distributions<sup>11,12</sup>, though C types outnumber S types beyond  $\sim 2.8$  AU. We ran a suite of dynamical simulations to investigate whether this giant planet migration scheme is consistent with the existence and structure of the asteroid belt. Because of the many unknowns in giant planet growth and early dynamical evolution,

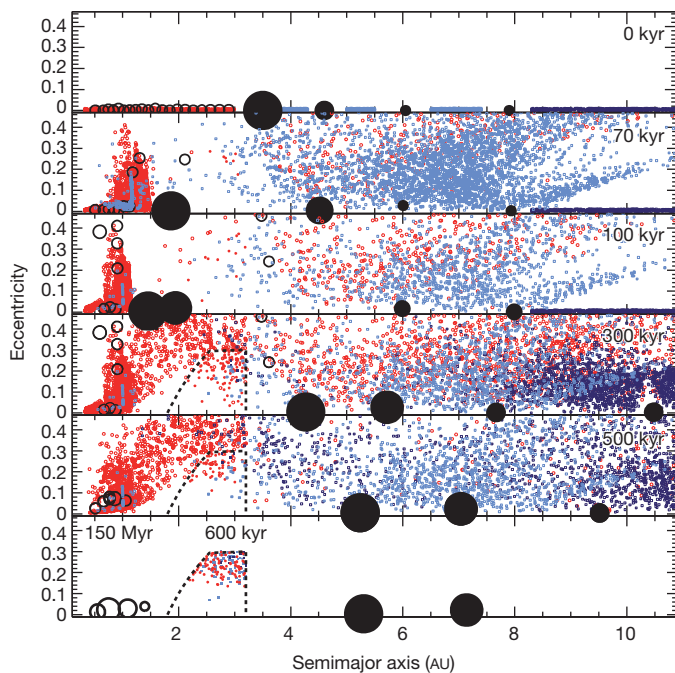
we present a simple scenario that reflects one plausible history for the giant planets (Fig. 1). We provide an exploration of parameter space (see Supplementary Information) that embraces a large range of possibilities and demonstrates the robustness of the results. In all simulations, we maintain the fundamental assumption that Jupiter tacked at 1.5 AU.

Figure 2 shows how the migration of the giant planets affects the small bodies. The disk interior to Jupiter has a mass 3.7 times that of the Earth ( $3.7M_{\oplus}$ ), equally distributed between planetary embryos (large)



**Figure 1 | The radial migration and mass growth imposed on the giant planets in the reference simulation.** **a**, Mass growth; **b**, semimajor axis. A fully-formed Jupiter starts at 3.5 AU, a location expected to be highly favourable for giant planet formation owing to the presence of the so-called snow line<sup>21</sup>. Saturn's  $30 M_{\oplus}$  core is initially at  $\sim 4.5$  AU and grows to  $60 M_{\oplus}$  as Jupiter migrates inward, over  $10^5$  years. Inward type-I migration of planetary cores is inhibited in disks with a realistic cooling timescale<sup>23–26</sup>; thus Saturn's core remains at 4.5 AU during this phase. Similarly, the cores of Uranus and Neptune begin at  $\sim 6$  and 8 AU and grow from  $5 M_{\oplus}$ , without migrating. Once Saturn reaches  $60 M_{\oplus}$  its inward migration begins<sup>25</sup>, and is much faster than that of the fully grown Jupiter<sup>27</sup>. Thus, on catching Jupiter, Saturn is trapped in the 2:3 resonance<sup>3</sup>. Here this happens when Jupiter is at 1.5 AU. The direction of migration is then reversed, and the giant planets migrate outward together. In passing, they capture Uranus and Neptune in resonance and these planets are then pushed outwards as well. Saturn, Uranus and Neptune reach their full mass at the end of the migration when Jupiter reaches 5.4 AU. The migration rate decreases exponentially as the gas disk dissipates. The final orbital configuration of the giant planets is consistent with their current orbital configuration when their later dynamical evolution is considered<sup>28,29</sup> (see Supplementary Information section 3 for extended discussion).

<sup>1</sup>Université de Nice – Sophia Antipolis, CNRS, Observatoire de la Côte d'Azur, BP 4229, 06304 Nice Cedex 4, France. <sup>2</sup>Department of Space Studies, Southwest Research Institute, 1050 Walnut Street, Suite 300, Boulder, Colorado 80302, USA. <sup>3</sup>Université de Bordeaux, Observatoire Aquitain des Sciences de l'Univers, 2 Rue de l'Observatoire, BP 89, F-33270 Floirac Cedex, France. <sup>4</sup>CNRS, UMR 5804, Laboratoire d'Astrophysique de Bordeaux, 2 Rue de l'Observatoire, BP 89, F-33270 Floirac Cedex, France. <sup>5</sup>Planetary Science Institute, 1700 East Fort Lowell, Suite 106, Tucson, Arizona 85719, USA. <sup>6</sup>NASA Goddard Space Flight Center, Code 693, Greenbelt, Maryland 20771, USA.



**Figure 2 | The evolution of the small-body populations during the growth and migration of the giant planets, as described in Fig. 1.** Jupiter, Saturn, Uranus and Neptune are represented by large black filled circles with evident inward-then-outward migration, and evident growth of Saturn, Uranus and Neptune. S-type planetesimals are represented by red dots, initially located between 0.3 and 3.0 AU. Planetary embryos are represented by large open circles scaled by  $M^{1/3}$  (but not in scale relative to the giant planets), where  $M$  is mass. The C-type planetesimals starting between the giant planets are shown as light blue dots, and the outer-disk planetesimals as dark blue dots, initially between 8.0 and 13.0 AU. For all planetesimals, filled dots are used if they are inside the main asteroid belt and smaller open dots otherwise. The approximate boundaries of the main belt are drawn with dashed curves. The bottom panel combines the end state of the giant planet migration simulation (including only those planetesimals that finish in the asteroid belt) with the results of simulations of inner disk material (semimajor axis  $a < 2$ ) evolved for 150 Myr (see Fig. 4), reproducing successful terrestrial planet simulations<sup>8</sup>.

and planetesimals (small), while the planetesimal population exterior to Jupiter is partitioned between inter-planetary belts and a trans-Neptunian disk (8–13 AU). The planetesimals from the inner disk are considered to be ‘S type’ and those from the outer regions ‘C type’. The computation of gas drag assumes 100-km-diameter planetesimals and uses a radial gas density profile taken directly from hydrodynamic simulations<sup>4</sup> (see Supplementary Information for details).

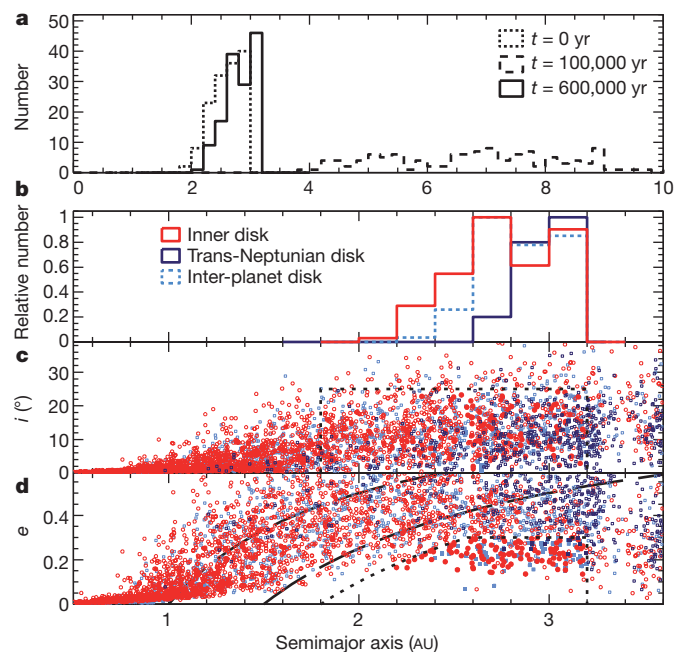
The inward migration of the giant planets shepherds much of the S-type material inward by resonant trapping, eccentricity excitation and gas drag. The mass of the disk inside 1 AU doubles, reaching  $\sim 2M_{\oplus}$ . This reshaped inner disk constitutes the initial condition for terrestrial planet formation. However, a fraction of the inner disk ( $\sim 14\%$ ) is scattered outward, ending up beyond 3 AU. During the subsequent outward migration of the giant planets, this scattered disk of S-type material is encountered again. Of this material, a small fraction ( $\sim 0.5\%$ ) is scattered inward and left decoupled from Jupiter in the asteroid belt region as the planets migrate away. The giant planets then encounter the material in the Jupiter–Neptune formation region, some of which ( $\sim 0.5\%$ ) is also scattered into the asteroid belt. Finally, the giant planets encounter the disk of material beyond Neptune (within 13 AU) of which only  $\sim 0.025\%$  reaches a final orbit in the asteroid belt. When the giant planets have finished their migration, the asteroid belt population is in place, whereas the terrestrial planets require an additional  $\sim 30$  Myr to complete their accretion.

The asteroid belt implanted in the simulations is composed of two separate populations: the S-type bodies originally from within 3.0 AU,

and the C types from between the giant planets and from 8.0 to 13.0 AU. The present-day asteroid belt consists of more than just S- and C-type asteroids, but this diversity is expected to result from compositional gradients within each parent population (Supplementary Information). There is a correlation between the initial and final locations of implanted asteroids (Fig. 3a). Thus, S-type objects dominate in the inner belt, while C-type objects dominate in the outer belt (Fig. 3b). Both types of asteroid share similar distributions of eccentricity and inclination (Fig. 3c, d). The present-day asteroid belt is expected to have had its eccentricities and inclinations reshuffled during the so-called late heavy bombardment (LHB)<sup>13,14</sup>; the final orbital distribution in our simulations matches the conditions required by LHB models.

Given the overall efficiency of implantation of  $\sim 0.07\%$ , our model yields  $\sim 1.3 \times 10^{-3} M_{\oplus}$  of S-type asteroids at the time of the dissipation of the solar nebula. In the subsequent 4.5 Gyr, this population will be depleted by 50–90% during the LHB event<sup>13,14</sup> and by a further factor of  $\sim 2$ –3 by chaotic diffusion<sup>15</sup>. The present-day asteroid belt is estimated to have a mass of  $6 \times 10^{-4} M_{\oplus}$ , of which 1/4 is S-type and 3/4 is C-type<sup>12</sup>. Thus our result is consistent within a factor of a few with the S-type portion of the asteroid belt.

The C-type share of the asteroid belt is determined by the total mass of planetesimals between the giant planets and between 8 and 13 AU, which are not known a priori. Requiring that the mass of implanted C-type material be three times that of the S-type, and given the implantation efficiencies reported above, this implies that the following



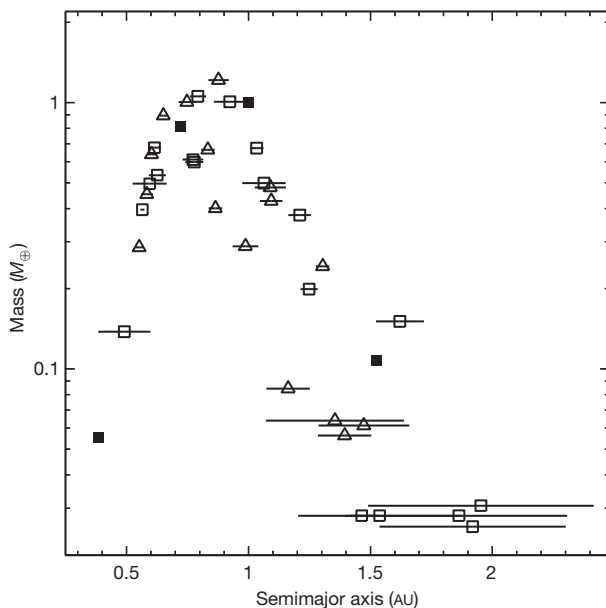
**Figure 3 | Distributions of 100-km planetesimals at the end of giant planet migration.** **a**, The semimajor axis distribution for the bodies of the inner disk that are implanted in the asteroid belt are plotted at three times: the beginning of the simulation (dotted histogram), at the end of inward planet migration (dashed) and at the end of outward migration (solid). There is a tendency for S-type planetesimals to be implanted near their original location. Thus, the outer edge of their final distribution is related to the original outer edge of the S-type disk, which in turn is related to the initial location of Jupiter. **b**, The final relative numbers of the S-type (red histogram), the inter-planet population (light blue) and the outer-disk (dark blue) planetesimals that are implanted in the asteroid belt are shown as a function of semimajor axis. The orbital inclination (**c**) and eccentricity (**d**) are plotted as a function of semimajor axis, with the same symbols used in Fig. 2. The dotted lines show the extent of the asteroid belt region for both inclination and eccentricity, and the dashed lines show the limits for perihelion less than 1.0 (left line) and 1.5 (right line). Most of the outer-disk material on planet-crossing orbits has high eccentricity, while many of the objects from between the giant planets were scattered earlier and therefore damped to lower-eccentricity planet-crossing orbits.



amount of material is left over from the giant planet accretion process:  $\sim 0.8 M_{\oplus}$  of material between the giant planets;  $\sim 16 M_{\oplus}$  of planetesimals from the 8.0–13 AU region; or some combination of the two.

Our simulations also found C-type material placed onto orbits crossing the still-forming terrestrial planets. For every C-type planetesimal from beyond 8 AU that was implanted in the outer asteroid belt, 11–28 C-type planetesimals ended up on high-eccentricity orbits that enter the terrestrial-planet-forming region (with perihelion  $q < 1.0$ –1.5 AU; see Fig. 3), and may represent a source of water for Earth<sup>16</sup>. For the Jupiter–Uranus region this ratio is 15–20, and for the Uranus–Neptune region it is 8–15. Thus, depending on which region dominated the implantation of C-type asteroids, we expect that  $(3–11) \times 10^{-2} M_{\oplus}$  of C-type material entered the terrestrial planet region. This exceeds by a factor of 6–22 the minimal mass required to bring the current amount of water to the Earth ( $\sim 5 \times 10^{-4} M_{\oplus}$ ; ref. 17), assuming that C-type planetesimals are 10% water by mass<sup>18</sup>.

We now consider the terrestrial planets. The migration of Jupiter creates a truncated inner disk matching initial conditions of previously successful simulations of terrestrial planet formation<sup>8</sup>, though there is a slight build-up of dynamically excited planetary embryos at 1.0 AU. Thus, we ran simulations of the accretion of the surviving objects for 150 Myr. Earth and Venus grow within the 0.7–1 AU annulus, accreting most of the mass, while Mars is formed from embryos scattered out beyond the edge of the truncated disk. Our final distribution of planet mass versus distance quantitatively reproduces the large mass ratio existing between Earth and Mars, and also matches quantitative metrics of orbital excitation (Fig. 4).



**Figure 4 | Results of the eight terrestrial planet simulations.** The mass versus semimajor axis of the synthetic planets (open symbols) is compared to the real planets (filled squares). The triangles refer to simulations starting with 40 embryos of  $\sim 0.051 M_{\oplus}$ , and squares to simulations from 80 embryos of  $\sim 0.026 M_{\oplus}$ . The horizontal error bars show the perihelion–aphelion excursion of each planet along their orbits. The initial planetesimal disk had an inner edge at 0.7 AU to replicate previous work<sup>8</sup>, and an outer edge at  $\sim 1.0$  AU owing to the truncation caused by the inward and outward migration of the giant planets as described in the text. Half of the original mass of the disk interior to Jupiter ( $1.85 M_{\oplus}$ ) was in  $\sim 727$  planetesimals. At the end of giant planet migration, the evolution of all objects inward of 2 AU was continued for 150 Myr, still accounting for the influence from Jupiter and Saturn. Collisions of embryos with each other and with planetesimals were assumed fully accretional. For this set of eight simulations, the average normalized angular momentum deficit<sup>30</sup> was  $0.0011 \pm 0.0006$ , as compared to 0.0018 for the current Solar System. Similarly, the radial mass concentration<sup>30</sup> was  $83.8 \pm 12.8$  as compared to 89.9 for the current Solar System.

Similar qualitative and quantitative results were found for a number of migration schemes, a range of migration and gas disk dissipation timescales, and a range of gas density and planetesimal sizes (all described in Supplementary Information). Our results represent a major shift in the understanding of the early evolution of the inner Solar System. In our scheme, C-type asteroids form between and beyond the giant planets, nearer to comets than to S-type asteroids. This could explain the substantial physical differences between S-type and C-type asteroids, and also the physical similarities between the latter and comets (as shown by the Stardust mission and micrometeorite samples<sup>19,20</sup>; see Supplementary Information for more on physical properties).

If Jupiter and Saturn have migrated substantially, then their birth region could have been closer to the estimated location of the snow line (the expected condensation front for water) at  $\sim 3$  AU (ref. 21), rather than out beyond 5 AU. Also, substantial migration is a point of similarity with observed extrasolar planetary systems, in which migration is seemingly ubiquitous—extrasolar giant planets are commonly found at  $\sim 1.5$  AU (refs 2, 22). However, a difference between our Solar System and the currently known extrasolar systems is that, according to our results, Jupiter ‘tacked’ at 1.5 AU and then migrated outward, owing to the presence of Saturn.

Received 1 September 2010; accepted 1 April 2011.

Published online 5 June 2011.

- Haisch, K. E. Jr, Lada, E. A. & Lada, C. J. Disk frequencies and lifetimes in young clusters. *Astrophys. J.* **553**, L153–L156 (2001).
- Armitage, P. J. Massive planet migration: theoretical predictions and comparison with observations. *Astrophys. J.* **665**, 1381–1390 (2007).
- Masset, F. & Snellgrove, M. Reversing type II migration: resonance trapping of a lighter giant protoplanet. *Mon. Not. R. Astron. Soc.* **320**, L55–L59 (2001).
- Morbidelli, A. & Crida, A. The dynamics of Jupiter and Saturn in the gaseous protoplanetary disk. *Icarus* **191**, 158–171 (2007).
- Pierens, A. & Nelson, R. P. Constraints on resonant-trapping for two planets embedded in a protoplanetary disc. *Astron. Astrophys.* **482**, 333–340 (2008).
- Kleine, T. et al. Hf–W chronology of the accretion and early evolution of asteroids and terrestrial planets. *Geochim. Cosmochim. Acta* **73**, 5150–5188 (2009).
- Wetherill, G. W. in *Protostars and Planets* (ed. Gehrels, T.) 565–598 (IAU Colloquium 52, International Astronomical Union, 1978).
- Hansen, B. M. S. Formation of the terrestrial planets from a narrow annulus. *Astrophys. J.* **703**, 1131–1140 (2009).
- Lin, D. N. C. & Papaloizou, J. On the tidal interaction between protoplanets and the protoplanetary disk. III — Orbital migration of protoplanets. *Astrophys. J.* **309**, 846–857 (1986).
- Crida, A. Minimum mass solar nebulae and planetary migration. *Astrophys. J.* **698**, 606–614 (2009).
- Gradie, J. & Tedesco, E. Compositional structure of the asteroid belt. *Science* **216**, 1405–1407 (1982).
- Mothé-Diniz, T., Carvano, J. M., Á. & Lazzaro, D. Distribution of taxonomic classes in the main belt of asteroids. *Icarus* **162**, 10–21 (2003).
- Gomes, R., Levison, H. F., Tsiganis, K. & Morbidelli, A. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* **435**, 466–469 (2005).
- Morbidelli, A., Brasser, R., Gomes, R., Levison, H. F. & Tsiganis, K. Evidence from the asteroid belt for a violent past evolution of Jupiter’s orbit. *Astron. J.* **140**, 1391–1401 (2010).
- Minton, D. A. & Malhotra, R. Dynamical erosion of the asteroid belt and implications for large impacts in the inner Solar System. *Icarus* **207**, 744–757 (2010).
- Morbidelli, A. et al. Source regions and time scales for the delivery of water to Earth. *Meteorit. Planet. Sci.* **35**, 1309–1320 (2000).
- Lécuyer, M., Gillet, P. & Robert, F. The hydrogen isotope composition of sea water and the global water cycle. *Chem. Geol.* **145**, 249–261 (1998).
- Abe, Y., Ohtani, E., Okuchi, T., Righter, K. & Drake, M. in *Origin of the Earth and Moon* (eds Canup, R. M. & Righter, K.) 413–433 (Univ. Arizona Press, 2000).
- Brownlee, D. et al. Comet 81P/Wild 2 under a microscope. *Science* **314**, 1711–1716 (2006).
- Gounelle, M. et al. in *The Solar System Beyond Neptune* (eds Barucci, M. A. et al.) 525–541 (Univ. Arizona Press, 2008).
- Ciesla, F. J. & Cuzzi, J. N. The evolution of the water distribution in a viscous protoplanetary disk. *Icarus* **181**, 178–204 (2006).
- Butler, R. P. et al. Catalog of nearby exoplanets. *Astrophys. J.* **646**, 505–522 (2006).
- Paardekooper, S. & Mellema, G. Halting type I planet migration in non-isothermal disks. *Astron. Astrophys.* **459**, L17–L20 (2006).
- Paardekooper, S. & Papaloizou, J. C. B. On disc protoplanet interactions in a non-barotropic disc with thermal diffusion. *Astron. Astrophys.* **485**, 877–895 (2008).
- Kley, W. & Crida, A. Migration of protoplanets in radiative discs. *Astron. Astrophys.* **487**, L9–L12 (2008).

26. Masset, F. S. & Casoli, J. On the horseshoe drag of a low-mass planet. II. Migration in adiabatic disks. *Astrophys. J.* **703**, 857–876 (2009).
27. Masset, F. S. & Papaloizou, J. C. B. Runaway migration and the formation of hot Jupiters. *Astrophys. J.* **588**, 494–508 (2003).
28. Morbidelli, A., Tsiganis, K., Crida, A., Levison, H. F. & Gomes, R. Dynamics of the giant planets of the Solar System in the gaseous protoplanetary disk and their relationship to the current orbital architecture. *Astron. J.* **134**, 1790–1798 (2007).
29. Batygin, K. & Brown, M. E. Early dynamical evolution of the Solar System: pinning down the initial conditions of the Nice model. *Astrophys. J.* **716**, 1323–1331 (2010).
30. Raymond, S. N., O'Brien, D. P., Morbidelli, A. & Kaib, N. A. Building the terrestrial planets: constrained accretion in the inner Solar System. *Icarus* **203**, 644–662 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** K.J.W. and A.M. were supported by the Helmholtz Alliances 'Planetary Evolution and Life' programme. S.N.R. and A.M.M. were supported by the

EPOV and PNP programmes of CNRS. D.P.O'B. was supported by the NASA PG&G programme. A.M.M. was also supported by the NASA post-doctoral programme and the Goddard Center for Astrobiology. We thank the Isaac Newton Institute DDP programme for hosting some of us at the initial stage of the project; we also thank J. Chambers for comments that improved the text. Computations were done on the CRIMSON Beowulf cluster at OCA.

**Author Contributions** K.J.W. managed the simulations and analysis and was the primary writer of the manuscript. A.M. initiated the project, updated and tested software, ran and analysed simulations, and wrote significant parts of the manuscript. S.N.R. helped initiate the project, advised on simulations and contributed substantially to the manuscript. D.P.O'B. helped initiate the project and assisted in writing. A.M.M. assisted in software updates and in writing.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to K.J.W. ([kwals@boulder.swri.edu](mailto:kwals@boulder.swri.edu)).



# Measurement of the internal state of a single atom without energy exchange

Jürgen Volz<sup>1</sup>, Roger Gehr<sup>1</sup>, Guilhem Dubois<sup>1†</sup>, Jérôme Estève<sup>1</sup> & Jakob Reichel<sup>1</sup>

A measurement necessarily changes the quantum state being measured, a phenomenon known as back-action. Real measurements, however, almost always cause a much stronger back-action than is required by the laws of quantum mechanics. Quantum non-demolition measurements have been devised<sup>1–6</sup> that keep the additional back-action entirely within observables other than the one being measured. However, this back-action on other observables often imposes its own constraints. In particular, free-space optical detection methods for single atoms and ions (such as the shelving technique<sup>7</sup>, a sensitive and well-developed method) inevitably require spontaneous scattering, even in the dispersive regime<sup>8</sup>. This causes irreversible energy exchange (heating), which is a limitation in atom-based quantum information processing, where it obviates straightforward reuse of the qubit. No such energy exchange is required by quantum mechanics<sup>9</sup>. Here we experimentally demonstrate optical detection of an atomic qubit with significantly less than one spontaneous scattering event. We measure the transmission and reflection of an optical cavity<sup>10–13</sup> containing the atom. In addition to the qubit detection itself, we quantitatively measure how much spontaneous scattering has occurred. This allows us to relate the information gained to the amount of spontaneous emission, and we obtain a detection error below 10 per cent while scattering less than 0.2 photons on average. Furthermore, we perform a quantum Zeno-type experiment to quantify the measurement back-action, and find that every incident photon leads to an almost complete state collapse. Together, these results constitute a full experimental characterization of a quantum measurement in the ‘energy exchange-free’ regime below a single spontaneous emission event. Besides its fundamental interest, this approach could significantly simplify proposed neutral-atom quantum computation schemes<sup>14</sup>, and may enable sensitive detection of molecules and atoms lacking closed transitions.

In the first step of any quantum measurement, the system to be measured becomes entangled with another quantum object (the ‘meter’), such as a photon field. For the case of a two-level system (a quantum bit, or qubit) with basis states  $|0\rangle$  and  $|1\rangle$ , the state  $(\alpha|0\rangle + \beta|1\rangle) \otimes |\Psi_{\text{in}}\rangle$  evolves into  $\alpha|0\rangle \otimes |\Psi_0\rangle + \beta|1\rangle \otimes |\Psi_1\rangle$ . The readout of the qubit then amounts to distinguishing the meter states  $|\Psi_0\rangle$  and  $|\Psi_1\rangle$ , which can only be achieved up to some error because they are generally non-orthogonal. The minimum possible detection error  $\varepsilon = (\varepsilon_0 + \varepsilon_1)/2$  is given by the Helstrom bound<sup>15</sup>

$$\varepsilon_{\text{H}} = \frac{1}{2} \left( 1 - \sqrt{1 - |\langle \Psi_0 | \Psi_1 \rangle|^2} \right) \quad (1)$$

where  $\varepsilon_0$  and  $\varepsilon_1$  are the probabilities of measuring the qubit in  $|1\rangle$  although it was in  $|0\rangle$  and vice versa, and we assume no prior knowledge of the qubit state. In the following, we consider the generic case where a qubit is probed by an incident coherent light pulse containing  $n$  photons on average. To a good approximation, the two final states  $|\Psi_0\rangle$  and  $|\Psi_1\rangle$  then also consist of coherent states. As an example, consider an ideal fluorescence measurement in which the dark state  $|0\rangle$  does not interact with the light, whereas the bright state  $|1\rangle$  scatters all photons. In this case,  $|\Psi_0\rangle = |0\rangle_{\text{S}}|n\rangle_{\text{T}}$  and  $|\Psi_1\rangle = |n\rangle_{\text{S}}|0\rangle_{\text{T}}$ ,

where  $|n\rangle$  is a coherent pulse containing  $n$  photons on average, and S and T refer to the scattered and transmitted light modes. Then,  $|\langle \Psi_0 | \Psi_1 \rangle|^2 = \exp(-2n)$ , and in the limit of large  $n$  one obtains  $\varepsilon_{\text{H}} \approx \exp(-2n)/4$ . More generally, in all schemes using coherent pulses (so that  $|\Psi_0\rangle$  and  $|\Psi_1\rangle$  are tensor products of coherent states each containing a photon number proportional to  $n$ ),  $|\langle \Psi_0 | \Psi_1 \rangle|^2 = \exp(-\zeta n)$  with some real value of  $\zeta$ . This exponential decrease of the minimum detection error with  $n$  naturally leads to a heuristic definition of the ‘knowledge’ of the atomic state as  $K \equiv \ln 2\varepsilon$ . The maximum knowledge ( $K_{\text{H}}$ ) that one can obtain from a measurement is then  $K_{\text{H}} = -\ln 2\varepsilon_{\text{H}}$ . We use the notation  $f(x) = -\ln(1 - \sqrt{1 - \exp(-x)})$ , which for large  $x$  simplifies to  $f(x) \approx x$ . Thus, for coherent pulse schemes,  $K_{\text{H}} = f(\zeta n)$  is the knowledge that the environment has obtained during the measurement, and constitutes an upper bound to the knowledge  $K_{\text{acc}}$  that the experimenter can actually access. In the case of the ideal fluorescence measurement, the maximum knowledge is  $f(2n) \approx 2n$ .

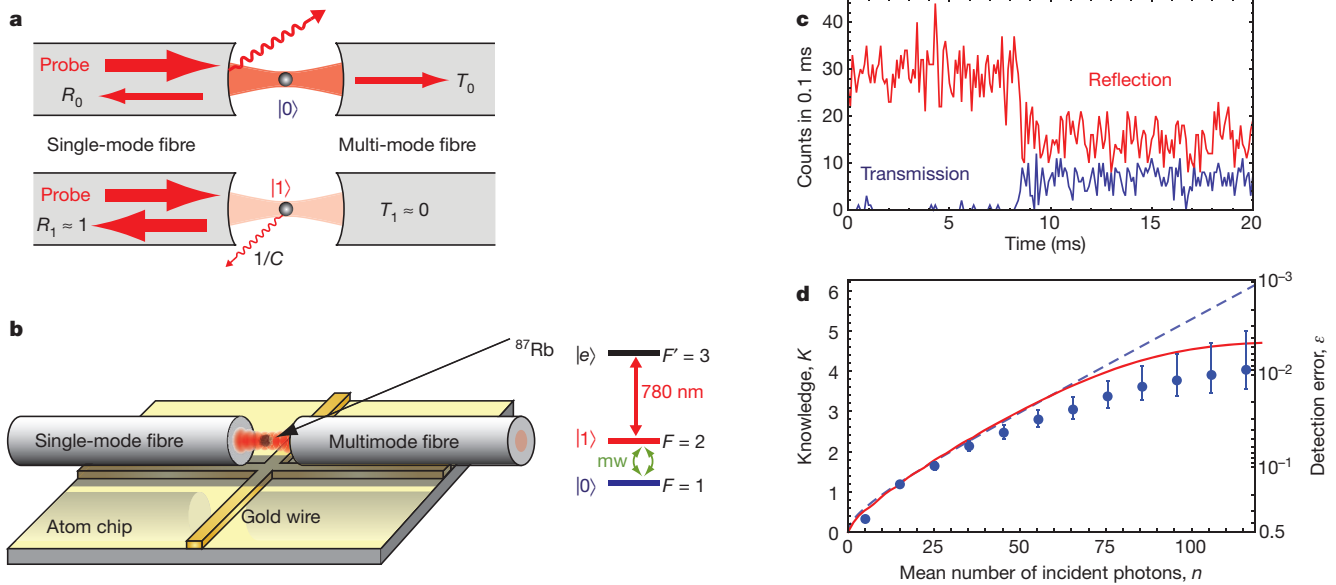
The measurement leads to a back-action on the atom. The final state of the qubit after the measurement is obtained by tracing over the meter<sup>16</sup>: the coherence of the qubit is reduced,  $\rho_{0,1} \rightarrow \langle \Psi_0 | \rho_{0,1} | \Psi_0 \rangle$ , where  $\rho$  is the qubit density matrix. This intrinsic back-action need not affect other degrees of freedom of the system: degrees of freedom that do not get entangled with the meter can be factored out and remain unaffected by the measurement. Most real measurements, however, cause a much larger back-action. In particular, fluorescence measurements are inevitably accompanied by spontaneous emission, which leads to heating and may pump the atom to an internal state outside the qubit basis. In the example of an ideal fluorescence detection, each incident photon is spontaneously scattered when the atom is in the bright state. Therefore, in terms of number of scattered photons,  $m$ , the maximum knowledge can be expressed as:

$$K_{\text{H}} = f(2m) \approx 2m \quad (2)$$

This bound in fact applies to all free-space measurement methods in which classical light sources are used in a single-pass configuration<sup>8</sup>: in all such methods, information gain is necessarily accompanied by energy exchange between the atom and the light. In particular, this includes dispersive measurements with far off-resonant light. Moreover, even state-of-the-art experiments typically fall short of this limit by several orders of magnitude (owing to limited collection efficiency), and require the scattering of a large number of photons to infer the qubit state<sup>17</sup>.

We overcome this limit by coupling the atomic qubit to a cavity in the strong-coupling regime,  $C = g^2/2\kappa\gamma \gg 1$ , where  $g$  describes the coherent atom–cavity coupling and  $\kappa$  ( $\gamma$ ) is the cavity (atomic) decay rate. The cavity is resonant with an optical transition of the  $|1\rangle$  state, and is probed by a resonant light pulse (Fig. 1a). An atom in the non-resonant state  $|0\rangle$  has a negligible effect on the cavity, and all photons from the incident mode are transmitted,  $|\Psi_0\rangle = |0\rangle_{\text{R}}|n\rangle_{\text{T}}$ . In contrast, an atom in the  $|1\rangle$  state detunes the cavity by more than its linewidth, so that almost all photons are reflected,  $|\Psi_1\rangle \approx |n\rangle_{\text{R}}|0\rangle_{\text{T}}$ . The states

<sup>1</sup>Laboratoire Kastler-Brossel, ENS, CNRS, Université Pierre et Marie Curie – Paris 6, 24 rue Lhomond, 75005 Paris, France. <sup>†</sup>Present address: Astron FIAMM, 35 rue Pasteur, 83210 La Fariède, France.



**Figure 1 | Cavity-assisted detection of an atomic qubit.** **a**, For an atom in the dark state  $|0\rangle$  (top), probe light is either transmitted, reflected or lost by mirror imperfections. For the bright state  $|1\rangle$  (bottom), most incident photons are reflected. In both cases, only a small fraction is scattered by the atom. **b**, Our cavity is formed by the coated end facets of two optical fibres. The qubit states ( $F=1, m_F=0$  and  $F=2, m_F=0$ ) can be coupled by a resonant microwave (mw) pulse. The cavity and the atomic transition  $|1\rangle \rightarrow |e\rangle$  are resonant with the  $\pi$ -polarized probe laser at 780 nm wavelength. **c**, Typical detection trace,

showing cavity transmission (blue) and reflection (red) for an atom initially in  $|1\rangle$  performing a quantum jump to  $|0\rangle$  owing to spontaneous emission. **d**, Data points, detection error and corresponding knowledge of the atomic state versus incident photon number  $n$  (error bars, 1 s.d.). Dashed line, theoretical prediction, taking into account our cavity imperfections (see text). We exclude the possibility of quantum jumps during the measurement, which explains the deviation for large  $n$ . Solid line, full simulation of our detection process, including quantum jumps.

$|\Psi_0\rangle$  and  $|\Psi_1\rangle$  thus have the same overlap as in the ideal fluorescence measurement, and  $K_H = f(2n)$  as before. Now, however, the atom sees a significant light intensity only when it is in the non-resonant state. Quantitatively, the  $|1\rangle$  state only scatters a fraction  $1/C$  of the incident photons<sup>18,19</sup>. Therefore, the maximum knowledge per scattered photon is  $C$  times larger than the free-space limit:

$$K_H = f(2Cm) \approx 2Cm \quad (3)$$

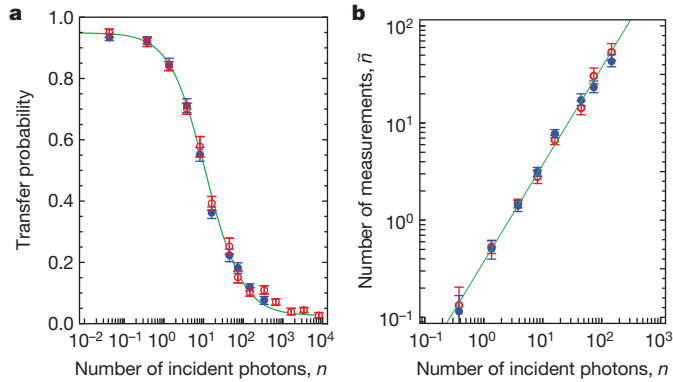
Furthermore, in contrast to fluorescence measurements,  $|\Psi_0\rangle$  and  $|\Psi_1\rangle$  are modes that are easily accessible experimentally. The atomic state can therefore be inferred with negligible spontaneous emission in a realistic experimental set-up.

Our implementation of this cavity readout scheme is shown in Fig. 1b. The key element is a fibre-based high-finesse cavity<sup>20,21</sup> ( $g = 2\pi \times 185(\pm 8)$  MHz,  $\kappa = 2\pi \times 53(\pm 0.5)$  MHz,  $\gamma = 2\pi \times 3$  MHz,  $C = 108(\pm 8)$ ) mounted on an atom chip. We prepare a Bose-Einstein condensate of  $^{87}\text{Rb}$  atoms, from which we load a single atom into an intracavity dipole trap<sup>22</sup>. Levels and transitions are shown in Fig. 1b. As shown earlier<sup>22</sup>, cavity transmission and reflection allow us to efficiently read out the atomic qubit state (Fig. 1c). Compared to the ideal situation described above, our system suffers from mirror losses and from the presence of the second TEM<sub>00</sub> cavity mode with orthogonal polarization detuned by 540 MHz, which has a non-negligible coupling to the atom (Methods). Losses reduce the empty-cavity transmission to  $T_0 = 0.13$  with associated reflection  $R_0 = 0.42$ . The presence of the second mode together with the effect of the hyperfine atomic structure degrades the extinction ratio of the transmission when a resonant atom is present. Instead of the ratio  $T_1/T_0 = 1/(4C^2)$  expected for a single-mode cavity coupled to a two-level atom, we measure  $T_1/T_0 = 2\%$  (Fig. 1c), where  $T_1 = 0.0024$  ( $R_1 \approx 1$ ) is the cavity transmission (reflection) with an atom in the resonant state  $|1\rangle$ . The two field states  $|\Psi_0\rangle$  and  $|\Psi_1\rangle$  now have additional components for the loss channels and for the outgoing modes coupled to the second cavity mode. This increases  $|\langle\Psi_0|\Psi_1\rangle|$ , leading to a reduced

$K_H = f(0.62n)$  (Methods). However, the intensity inside the cavity is also reduced by the mirror loss and therefore the expected knowledge in terms of scattered photons is still much higher than  $m$  (Methods).

Knowledge of the atomic state carried by photons lost at the mirrors is not accessible to the experimenter, reducing the available knowledge to  $f(0.23n)$ . Furthermore, photon counting in the reflected and transmitted modes is not an optimal strategy to distinguish the two states  $|\Psi_0\rangle$  and  $|\Psi_1\rangle$ . The associated detection error  $\epsilon_d$  is given by the overlap of the two probability distributions of counts detected when the atom is in  $|0\rangle$  or  $|1\rangle$  (see Methods). As predicted by the Chernoff bound<sup>23</sup>, it decreases exponentially for large  $n$  with rate  $\xi$ , where  $\xi$  can be calculated from the reflection and transmission coefficients of the cavity and the efficiency of the photon detectors (Methods). Therefore, in the large- $n$  limit, the knowledge that can be experimentally accessed,  $K_{\text{acc}} = -\ln 2\epsilon_d$ , follows the function  $f(x)$  introduced above, with  $x = \xi n$ . We checked numerically that  $K_{\text{acc}} \approx f(\xi n)$  is also a valid approximation for small  $n$ . Taking into account finite photon detection efficiencies (47% in transmission, 31% in reflection), we expect our detection method to yield  $K_{\text{acc}} = f(4.6 \times 10^{-2}n)$  (which would increase to  $f(0.11n)$  using perfect detectors). To verify this prediction, we prepare the atom in either of the qubit states  $|0\rangle$  and  $|1\rangle$  and measure the corresponding detection errors<sup>22</sup>. For  $n < 40$ , the measurement is in good agreement with the prediction (Fig. 1d). For larger  $n$ , non-resonant excitation leads to a small probability of pumping the qubit from its initial state to the other hyperfine state during the measurement<sup>22</sup>, thereby reducing the accessible knowledge in the experiment.

Although we only detect a fraction of the incident photons and therefore of  $K_H$ , we can still quantify  $K_H$  by its back-action. We perform the following quantum Zeno<sup>24</sup> experiment: an atom is prepared in state  $|0\rangle$  or  $|1\rangle$  (Methods) and a microwave  $\pi$ -pulse applied on the  $|0\rangle \leftrightarrow |1\rangle$  resonance. During the  $\pi$ -pulse of duration  $\tau$ , we apply detection light with a variable average photon number  $n$ . The incident light measures the atomic state and thereby prevents the Rabi oscillation, as shown in Fig. 2a. We model this system as a coherently driven qubit

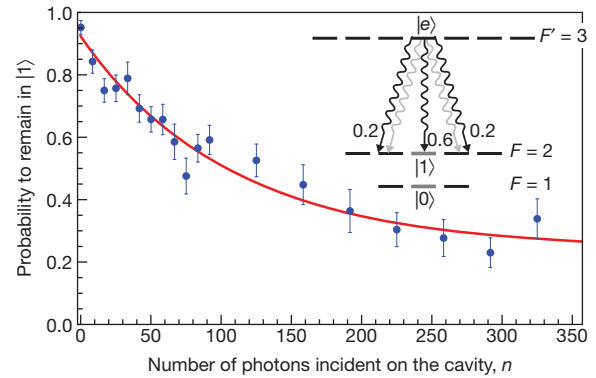


**Figure 2 | Back-action measurement using the quantum Zeno effect.** A microwave  $\pi$ -pulse (duration 8.8  $\mu$ s) is applied to an atom in  $|1\rangle$  (blue filled circles) or  $|0\rangle$  (red open circles) in the presence of measurement light. **a**, Data points, transfer efficiency versus number of photons incident on the cavity during the pulse (error bars, 1 s.d.). **b**, The average number of projective measurements  $\tilde{n}$  that we deduce for each data point from our model; the solid line is a linear fit  $\tilde{n} = a_0 n$ , yielding  $a_0 = 0.37 \pm 0.02$ . The solid line in **a** shows the prediction of our theoretical model supposing this linear relation and this value of  $a_0$ .

undergoing on average  $\tilde{n}$  projective measurements that are randomly distributed over  $\tau$ , and solve the corresponding Bloch equations. We include technical imperfections (preparation, detection and pulse errors) which limit the maximal (minimal) transfer probability to 0.95 (0.02). From the measured transfer efficiency (Fig. 2a), we infer  $\tilde{n}$  for each mean number of incident photons  $n$  and, as expected, we observe a linear relationship  $\tilde{n} = (0.37 \pm 0.02)n$  (Fig. 2b). To see how the maximum knowledge  $K_H$  is related to  $\tilde{n}$ , we consider the evolution of the qubit in the absence of the microwave pulse. The Bloch equations predict a reduction of coherence by  $\exp(-\tilde{n})$ , whereas the equivalent description introduced before (partial trace over the ‘meter’) predicts  $\langle \Psi_1 | \Psi_0 \rangle = \exp(-\tilde{n})$ , and  $K_H = f(2\tilde{n})$ . The Zeno measurement therefore yields  $K_H = f(0.74 \pm 0.04)n$ , in reasonable agreement with the value expected from photonic mode overlap,  $K_H = f(0.62n)$ . This shows that every single photon incident on the cavity leads to a significant state collapse, reducing the atomic coherence by a factor of 0.7.

To relate the maximum knowledge  $K_H$  and the accessible knowledge  $K_{acc}$  to the number of scattered photons  $m$ , we measure  $m$  as a function of  $n$ . Rather than attempting direct detection of the spontaneous photons (which would be inefficient and difficult to calibrate), we take advantage of the fact that each spontaneous scattering event of the  $|1\rangle = |F=2, m=0\rangle$  state has a known probability of depumping to other states  $|F=2, m \neq 0\rangle$ . The scattering rate of the off-resonant state  $|0\rangle$  is three orders of magnitude smaller and can be neglected. We prepare the atom in  $|1\rangle$  and turn on detection light for a variable time. Then we apply a microwave  $\pi$ -pulse on the  $|1\rangle \leftrightarrow |0\rangle$  transition, and finally we determine whether the atom has been transferred to  $|0\rangle$ . The microwave pulse has no effect on initial states  $|F=2, m \neq 0\rangle$ , so that the probability of the atom transferring to  $|0\rangle$  is equal to the fraction of population remaining in  $|1\rangle$  after the detection light pulse. We find that this survival probability decays exponentially with the number of incident photons, with an initial rate  $\nu = 1/(142 \pm 25)$  (Fig. 3). Because of the probability of decaying back to  $|1\rangle$ , the actual spontaneous emission rate is larger than this depumping rate. Correcting for this effect (Methods), we obtain  $m/n = 1/(118 \pm 20)$ . This is compatible with the theoretical prediction  $m/n = 1/83$  for our particular atom–cavity system, where the second cavity mode increases the spontaneous emission rate (Methods). A still better value,  $m/n = \sqrt{T_0}/C$ , can be expected for a single-mode cavity.

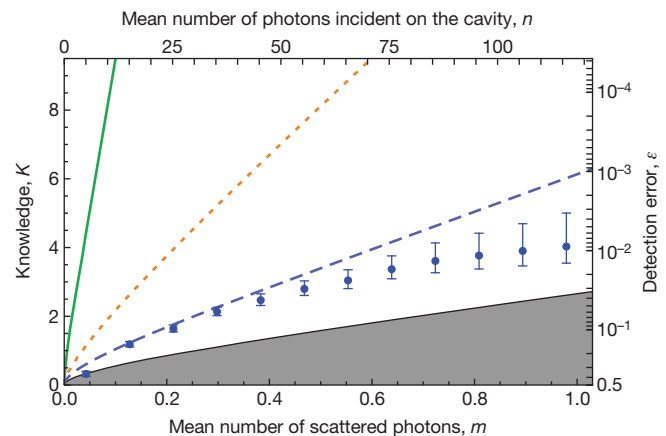
We can now express  $K_H$  and  $K_{acc}$  in terms of scattered photons (Fig. 4). In the regime  $m \ll 1$ , where the detection efficiency is not



**Figure 3 | Spontaneous emission during detection.** Data points, the measured probability that the atom remains in  $|1\rangle$  during detection versus the number of incident photons  $n$  (error bars, 1 s.d.). Solid line, a fit of an exponential decay to a steady-state population of  $0.27 \pm 0.05$  with initial rate  $\nu = 1/(142 \pm 25)$ . Inset, the two decay processes depleting  $|1\rangle$ . Detection light excites the state  $|e\rangle$ , which can decay into free space (black) with rate  $\Gamma$  or into the second cavity mode (light grey) with rate  $\Gamma_p$  (Methods). Correcting for decay back to  $|1\rangle$ , we obtain the number of scattered photons  $m = n/(118 \pm 20)$ .

limited by depumping, we find that our experiment extracts  $K_H = f(87 \pm 17)m$ , of which  $K_{acc} = f(5.4 \pm 0.9)m$  is actually accessed. In spite of experimental imperfections, this knowledge gain is a factor of 2.7 higher than possible in an ideal fluorescence measurement and two orders of magnitude larger than in state-of-the-art experiments<sup>17</sup>. Because 118 photons on average can be sent onto the cavity before one scattering event occurs and each photon performs a strong measurement, a large amount of information on the atomic state can be obtained with negligible scattering. In this sense, one can say that the photons measure the atom without entering the cavity.

Note that our experiment is still limited by cavity imperfections that should be straightforward to improve. The closely spaced second cavity mode is a result of birefringence; experience with macroscopic cavities suggests that it can be either moved further away or made degenerate in future cavities. Cavity losses can be further reduced by at least a factor of four<sup>21</sup> by using state-of-the-art mirror coatings in an otherwise identical fibre cavity. Assuming these conditions and



**Figure 4 | Detection error and knowledge versus number of scattered photons.** The grey area is the range accessible to free-space detection schemes; this limit is overcome using a cavity. Green solid line, maximum knowledge  $K_H$  extracted by the cavity measurement, deduced from the data in Fig. 2. Orange dashed line, accessible knowledge using our cavity with perfect photon counters to detect reflected and transmitted photons. Blue dashed line, accessible information  $K_{acc}$  using the detection efficiency of our experiment. Filled blue circles, knowledge actually obtained from the experiment (error bars, 1 s.d.). This knowledge is above the free-space limit despite experimental imperfections.



detector efficiencies of 70%, the accessible knowledge would be  $K_{\text{acc}} \approx f(110m)$ .

For our cavity parameters, heating mechanisms other than scattering are expected to be negligible<sup>25</sup>. Furthermore, the remaining scattering only leads to a small probability of the atom changing its vibrational state during detection, because of the Lamb–Dicke effect in our strong dipole trap. We estimate this probability to be two orders of magnitude smaller than in state-of-the-art fluorescence measurements for the same knowledge gain (in our case we can access  $K_{\text{acc}} \approx 70$  before the creation of a phonon). This allows detection of atomic qubits while staying in the ground state of the trap, thereby removing the necessity of recoiling after read-out and drastically improving the cycling time of atom-based quantum computing schemes. Furthermore, the cavity readout scheme disposes of the requirement for closed transitions in state readout, thereby opening up ways to detect single cold molecules<sup>26</sup>.

## METHODS SUMMARY

**Preparation of the qubit states.** We initially extract a single atom in the  $F = 2$  hyperfine ground state from a Bose–Einstein condensate<sup>22</sup>. Then we apply a microwave  $\pi$ -pulse on the qubit transition  $|0\rangle \leftrightarrow |1\rangle$ , followed by a short detection light pulse. If the atom initially is in  $|1\rangle$ , the  $\pi$ -pulse transfers it to  $|0\rangle$ , leading to high cavity transmission. Otherwise, the atom remains in  $F = 2$ , leading to low transmission. In this case, spontaneous scattering due to the readout pulse leads to a redistribution in the  $F = 2$  multiplet. We repeat the procedure until we detect high transmission, signalling an atom in  $|0\rangle$ . To prepare the atom in  $|1\rangle$  we apply an additional  $\pi$ -pulse.

**Extracted and accessible knowledge from a lossy cavity.** For each of the two orthogonally polarized cavity modes, the transmitted, reflected and lost field can each be approximated by a coherent state with amplitude  $\alpha_i \sqrt{n}$ , where  $i$  denotes the qubit state and  $n$  is the incident photon number. The overlap between the two possible light states  $|\Psi_0\rangle$  and  $|\Psi_1\rangle$  is then  $\exp(-\zeta n)$ , where  $\zeta = \sum_i |\alpha_0 - \alpha_i|^2$ , and the sum goes over the reflected, transmitted and lost modes. This yields the maximum knowledge  $f(\zeta n)$ , where  $\zeta = 0.62$  for our cavity parameters. Using counters to detect the transmitted and reflected photons, the accessible knowledge is approximately  $f(\xi n)$  with  $\xi = 4.6 \times 10^{-2}$  (Methods).

**Determination of spontaneous scattering.** The detection light populates the excited state  $|F' = 3, m_F = 0\rangle$ , which has two decay channels: spontaneous emission into free space with rate  $\Gamma$ , and into the second orthogonally polarized TEM<sub>00</sub> cavity mode (detuned by 540 MHz; ref. 22) with a Purcell-enhanced rate  $\Gamma_p$ . From a numerical simulation of the atom–cavity master equation, we obtain the ratio of the two decay channels,  $\Gamma_p/\Gamma = 2.6$ . Because of the probability of decaying back into the original state  $|1\rangle$ , the total spontaneous emission rate  $\Gamma + \Gamma_p$  is higher than the measured decay constant of state  $|1\rangle$ . We correct for this small effect by using the known transition strengths, and obtain  $n/m = v^{-1} \times (\Gamma_p + 2\Gamma/5)/(\Gamma_p + \Gamma)$ .

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 21 January; accepted 19 May 2011.

1. Braginsky, V. B. & Khalili, F. Y. *Quantum Measurement* (Cambridge Univ. Press, 1992).

2. Grangier, P., Levenson, J. A. & Poizat, J.-P. Quantum non-demolition measurements in optics. *Nature* **396**, 537–542 (1998).
3. Nogues, G. *et al.* Seeing a photon without destroying it. *Nature* **400**, 239–242 (1999).
4. Maioli, P. *et al.* Nondestructive Rydberg atom counting with mesoscopic fields in a cavity. *Phys. Rev. Lett.* **94**, 113601 (2005).
5. Hume, D. B., Rosenband, T. & Wineland, D. J. High-fidelity adaptive qubit detection through repetitive quantum nondemolition measurements. *Phys. Rev. Lett.* **99**, 120502 (2007).
6. Lupaşcu, A. *et al.* Quantum non-demolition measurement of a superconducting two-level system. *Nature Phys.* **3**, 119–125 (2007).
7. Leibfried, D., Blatt, R., Monroe, C. & Wineland, D. Quantum dynamics of single trapped ions. *Rev. Mod. Phys.* **75**, 281–324 (2003).
8. Hope, J. J. & Close, J. D. General limit to nondestructive optical detection of atoms. *Phys. Rev. A* **71**, 043822 (2005).
9. Kwiat, P., Weinfurter, H., Herzog, T., Zeilinger, A. & Kasevich, M. A. Interaction-free measurement. *Phys. Rev. Lett.* **74**, 4763–4766 (1995).
10. Boozer, A. D., Boca, A., Miller, R., Northup, T. E. & Kimble, H. J. Cooling to the ground state of axial motion for one atom strongly coupled to an optical cavity. *Phys. Rev. Lett.* **97**, 083602 (2006).
11. Puppe, T. *et al.* Trapping and observing single atoms in a blue-detuned intracavity dipole trap. *Phys. Rev. Lett.* **99**, 013002 (2007).
12. Khudaverdyan, M. *et al.* Quantum jumps and spin dynamics of interacting atoms in a strongly coupled atom-cavity system. *Phys. Rev. Lett.* **103**, 123006 (2009).
13. Bochmann, J. *et al.* Lossless state detection of single neutral atoms. *Phys. Rev. Lett.* **104**, 203601 (2010).
14. Ladd, T. D. *et al.* Quantum computers. *Nature* **464**, 45–53 (2010).
15. Helstrom, C. W. *Mathematics in Science and Engineering* Vol. 123, *Quantum Detection and Estimation Theory* (Academic, 1976).
16. Haroche, S. & Raimond, J.-M. *Exploring the Quantum* (Oxford Univ. Press, 2006).
17. Gerber, S. *et al.* Quantum interference from remotely trapped ions. *N. J. Phys.* **11**, 013032 (2009).
18. Lugiato, L. A. in *Progress in Optics* Vol. XXI (ed. Wolf, E.) 69–216 (Elsevier Science, 1984).
19. Hechenblaikner, G., Gangl, M., Horak, P. & Ritsch, H. Cooling an atom in a weakly driven high-Q cavity. *Phys. Rev. A* **58**, 3030–3042 (1998).
20. Colombe, Y. *et al.* Strong atom-field coupling for Bose–Einstein condensates in an optical cavity on a chip. *Nature* **450**, 272–276 (2007).
21. Hunger, D. *et al.* A fiber Fabry–Perot cavity with high finesse. *N. J. Phys.* **12**, 065038 (2010).
22. Gehr, R. *et al.* Cavity-based single atom preparation and high-fidelity hyperfine state readout. *Phys. Rev. Lett.* **104**, 203602 (2010).
23. Chernoff, H. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.* **23**, 493–507 (1952).
24. Itano, W. M. Perspectives on the quantum Zeno paradox. Preprint at (<http://arxiv.org/abs/quant-ph/0612187v1>) (2006).
25. Domokos, P. & Ritsch, H. Mechanical effects of light in optical resonators. *J. Opt. Soc. Am. B* **20**, 1098–1130 (2003).
26. Jones, K. M., Tiesinga, E., Lett, P. D. & Julienne, P. S. Ultracold photoassociation spectroscopy: long-range molecules and atomic scattering. *Rev. Mod. Phys.* **78**, 483–535 (2006).

**Acknowledgements** This work was funded in part by the AQUITE Integrated Project of the EU (grant no. 247687), by the Institut Francilien pour la Recherche sur les Atomes Froids (IFRAF), and by the EURYI grant ‘Integrated Quantum Devices’.

**Author Contributions** J.V., R.G. and G.D. performed the experiment. All authors contributed to data analysis and interpretation, as well as to the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.R. ([jakob.reichel@ens.fr](mailto:jakob.reichel@ens.fr)).

## METHODS

**Preparation of the qubit states.** We initially extract a single atom in the  $F=2$  hyperfine ground state and unknown Zeeman state from a Bose–Einstein condensate<sup>22</sup>. We then apply a microwave  $\pi$ -pulse on the qubit transition  $|0\rangle \leftrightarrow |1\rangle$ , followed by a short detection light pulse. If and only if the atom initially is in  $|1\rangle$ , the  $\pi$ -pulse transfers it to  $|0\rangle$ , leading to high cavity transmission. If it is in a different Zeeman sublevel, it remains in  $F=2$ , leading to low transmission. In this case, spontaneous scattering due to the read out pulse leads to a redistribution in the  $F=2$  multiplet. We repeat the procedure until we detect high transmission, signalling an atom in  $|0\rangle$ . If we want to prepare the atom in  $|1\rangle$ , we apply an additional  $\pi$ -pulse.

**Extracted and accessible knowledge from an imperfect cavity.** Assuming a coherent state with amplitude  $\sqrt{n}$  incident onto the cavity, a coherent field builds up in the cavity populating the main (resonant) and the orthogonally polarized, detuned TEM00 mode. Their amplitudes depend on the qubit state and each decays via three channels: transmission, reflection and losses at the mirrors. Thus the outgoing field can be approximated by the tensor product of six coherent fields with amplitudes  $\alpha_i \sqrt{n}$ . Here the subscript  $i \in \{0, 1\}$  denotes the qubit state, while  $\alpha \in \{t_m, r_m, l_m, t_d, r_d, l_d\}$  identifies the outgoing mode,  $m$  and  $d$  respectively designating the main and detuned mode. The overlap between the two possible light states  $|\Psi_0\rangle$  and  $|\Psi_1\rangle$  is then  $\exp(-\zeta n)$  where  $\zeta = \sum_i |\alpha_{0i} - \alpha_{1i}|^2$ , leading to a maximum knowledge  $f(\zeta n)$ . In our case, the atom in state  $|1\rangle$  is resonant with cavity and light field whereas the state  $|0\rangle$  is far detuned. Under these circumstances, the phase shift between the states  $|\Psi_1\rangle$  and  $|\Psi_0\rangle$  can be neglected and the amplitude coupling factors  $\{\alpha_i\}$  can be considered real. The power coupling factors in per cent are  $\{\alpha_0^2\} = \{12.7, 41.4, 45.9, 0, 0, 0\}$  for an atom in  $|0\rangle$  and  $\{\alpha_1^2\} = \{0.1, 99, 0.4, 0.1, 0.1, 0.4\}$  for an atom in  $|1\rangle$ . From these values, we deduce  $\zeta = 0.62$ .

Using counters to detect the transmitted and reflected intensities, the detection error is minimized by using a maximum likelihood estimator (thresholding). The minimal error is  $\varepsilon_D = (1 - \|P_0 - P_1\|_1)/2$ , where  $P_0$  and  $P_1$  are the probability distributions of detected counts if the qubit is in  $|0\rangle$  or  $|1\rangle$ , and the distance

$\|P_0 - P_1\|_1$  is defined as  $\sum_x |P_0(x) - P_1(x)|/2$ , where the sum goes over all possible detection events. In our case, the two distributions are the products of two Poisson distributions and we have numerically observed that their distance is well approximated by  $\sqrt{1-Q}$ , where  $Q$  is the Chernoff coefficient  $Q = \min_{0 \leq s \leq 1} [\sum_x P_0^s(x) P_1^{1-s}(x)]$ , which can be calculated to be  $Q = \exp(-\zeta n)$ , where  $\zeta = -\min_{0 \leq s \leq 1} [T_0^s T_1^{1-s} + R_0^s R_1^{1-s} - s(T_0 + R_0) - (1-s)(T_1 + R_1)]$ . The accessible knowledge is then given by the same expression  $f(\zeta n)$  as the maximum knowledge derived from the Helstrom error bound (but of course with a different  $\zeta$ ), allowing direct comparisons between the two. In our regime of parameters, the minimum is reached for  $s \approx 0.5$  leading to  $\zeta = (T_0 + T_1 + R_0 + R_1)/2 - \sqrt{T_0 T_1} - \sqrt{R_0 R_1} = 0.11$ . Taking into account the finite detector efficiencies, the same calculation leads to  $\zeta = 4.6 \times 10^{-2}$ .

**Determination of the spontaneous scattering rate.** The detection light populates the excited state  $|F'=3, m_F=0\rangle$ , which can decay via two channels: spontaneous emission into free space with rate  $\Gamma$ , and into the second orthogonally polarized TEM00 cavity mode which is detuned by 540 MHz (ref. 22) with a Purcell-enhanced rate  $\Gamma_p$  (Fig. 3). Decay into the original (pumped) cavity mode is not considered because it constitutes a coherent process which does not change the atomic state. The decay into the second cavity mode always leads to a change of the atomic Zeeman state, whereas for the decay into free space the atom has a probability of 3/5 (given by the transition strengths) of ending up in the original Zeeman ground state. Therefore, the total spontaneous emission rate  $\Gamma + \Gamma_p$  is higher than the measured decay constant of state  $|1\rangle$  (Fig. 3). To correct for this small effect, we have to know the relative probability of the two decay channels. The effect of the second cavity mode is too strong to be treated as a perturbation, therefore we numerically solve the complete master equation (including all ground and excited state Zeeman levels as well as the two cavity modes). From this solution, we obtain the ratio of the two decay channels of  $\Gamma_p/\Gamma = 2.6$ , which depends only weakly on experimental parameters. Using this value, we obtain for the spontaneous scattering rate  $n/m = v^{-1} \times (\Gamma_p + 2\Gamma/5)/(\Gamma_p + \Gamma)$ . This ratio is 3.6 times smaller than the value  $C/\sqrt{T_0}$  for a single-mode cavity.

# Increased soil emissions of potent greenhouse gases under increased atmospheric CO<sub>2</sub>

Kees Jan van Groenigen<sup>1,2,3</sup>, Craig W. Osenberg<sup>4</sup> & Bruce A. Hungate<sup>1,2</sup>

Increasing concentrations of atmospheric carbon dioxide (CO<sub>2</sub>) can affect biotic and abiotic conditions in soil, such as microbial activity and water content<sup>1,2</sup>. In turn, these changes might be expected to alter the production and consumption of the important greenhouse gases nitrous oxide (N<sub>2</sub>O) and methane (CH<sub>4</sub>) (refs 2, 3). However, studies on fluxes of N<sub>2</sub>O and CH<sub>4</sub> from soil under increased atmospheric CO<sub>2</sub> have not been quantitatively synthesized. Here we show, using meta-analysis, that increased CO<sub>2</sub> (ranging from 463 to 780 parts per million by volume) stimulates both N<sub>2</sub>O emissions from upland soils and CH<sub>4</sub> emissions from rice paddies and natural wetlands. Because enhanced greenhouse-gas emissions add to the radiative forcing of terrestrial ecosystems, these emissions are expected to negate at least 16.6 per cent of the climate change mitigation potential previously predicted from an increase in the terrestrial carbon sink under increased atmospheric CO<sub>2</sub> concentrations<sup>4</sup>. Our results therefore suggest that the capacity of land ecosystems to slow climate warming has been overestimated.

By burning fossil fuels, cutting down forests and changing land use in other ways, humans are rapidly increasing the amount of CO<sub>2</sub> in the atmosphere and warming the planet<sup>5</sup>. Plant growth is known to increase after an abrupt surge in CO<sub>2</sub> levels<sup>6</sup>. Because stimulated assimilation of carbon by plants can increase soil carbon input and soil carbon storage, terrestrial ecosystems could help to reduce the increase in atmospheric CO<sub>2</sub> and thereby slow climate change<sup>7</sup>. However, the radiative forcing of land ecosystems is not determined by their uptake and release of CO<sub>2</sub> alone; increased CO<sub>2</sub> can also alter soil emissions of N<sub>2</sub>O and CH<sub>4</sub> (ref. 2). Although both of these gases occur in far lower atmospheric concentrations than does CO<sub>2</sub>, their global warming potentials are much higher: 298 times higher for N<sub>2</sub>O and 25 times higher for CH<sub>4</sub> (ref. 5). Agricultural soils are the main source of human-induced N<sub>2</sub>O emissions<sup>8</sup>. Soils under natural vegetation produce roughly the same amount of N<sub>2</sub>O as all anthropogenic sources combined<sup>8</sup>. Wetlands, including rice paddies, contribute 32–53% to the global emissions of CH<sub>4</sub> (ref. 8). Upland soils, on the other hand, act as a sink for atmospheric CH<sub>4</sub> through oxidation by methanotrophic bacteria<sup>9</sup>. Thus, changes in N<sub>2</sub>O and CH<sub>4</sub> fluxes could greatly alter how terrestrial ecosystems influence climate<sup>10</sup>.

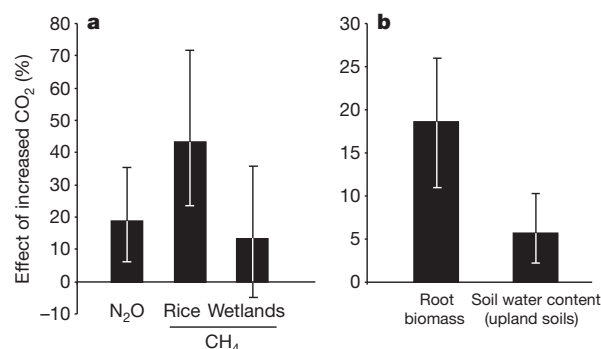
Studies of greenhouse-gas (GHG) emissions span a variety of ecosystem types, and vary in experimental design and results, making it difficult to determine their global response to increased CO<sub>2</sub> from individual experiments. A quantitative synthesis of results across multiple studies can overcome this problem. Therefore, we used meta-analysis<sup>11</sup> to summarize the effect of atmospheric CO<sub>2</sub> enrichment on fluxes of CH<sub>4</sub> and N<sub>2</sub>O from soil, using 152 observations from 49 published studies (see Supplementary Table 1, Supplementary Data 1 and 2, Supplementary Notes 1). We also summarized the effect of increased CO<sub>2</sub> on possible drivers of altered CH<sub>4</sub> and N<sub>2</sub>O fluxes, using standing root biomass and soil water content from the studies in which the observations on N<sub>2</sub>O and CH<sub>4</sub> fluxes were collected (Supplementary Data 3 and 4). All observations were analysed using three different weighting functions (see Methods). As CH<sub>4</sub> and N<sub>2</sub>O

emissions were not correlated with the concentration of CO<sub>2</sub> used for enrichment (Methods), we treat ‘increased CO<sub>2</sub>’ as a category.

Overall, increased concentrations of atmospheric CO<sub>2</sub> stimulated emissions of N<sub>2</sub>O by 18.8% (Fig. 1a). This positive response was significant for studies receiving little or no fertilizer, for non-pot studies and for studies on natural vegetation—that is, studies that most closely resembled real-world conditions (Supplementary Table 2). Increased CO<sub>2</sub> stimulated CH<sub>4</sub> emissions in wetlands by 13.2% (Fig. 1a, Supplementary Table 3). In rice paddies, increased CO<sub>2</sub> stimulated CH<sub>4</sub> emissions by 43.4% (Fig. 1a, Supplementary Table 4). In upland systems, increased CO<sub>2</sub> caused on average a small and insignificant net uptake of CH<sub>4</sub> (Supplementary Table 5).

To compare the relative importance of changed GHG fluxes in uplands, wetlands and rice paddies, we expressed the absolute effect of increased CO<sub>2</sub> on CH<sub>4</sub> and N<sub>2</sub>O fluxes from these ecosystem types (Supplementary Tables 5–8) scaled by their respective total land area. For upland soils, we distinguished fertilized agricultural ecosystems and ecosystems receiving little or no fertilizer. Our estimates of total GHG fluxes under ambient (that is, present-day) CO<sub>2</sub> conditions correspond well to independent global syntheses of modern GHG fluxes (Supplementary Table 9), supporting our scaling approach.

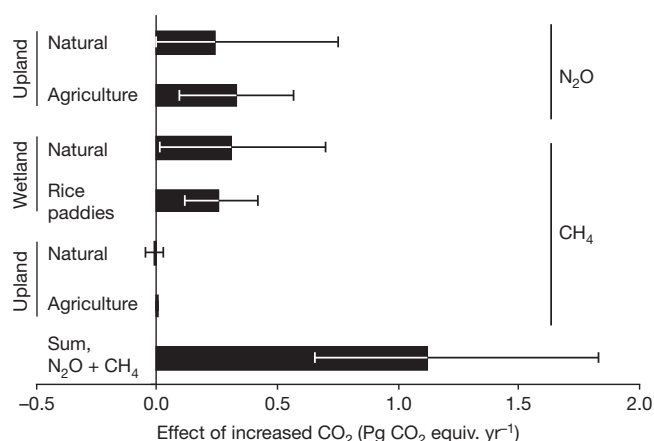
The estimated stimulation by increased CO<sub>2</sub> of total soil N<sub>2</sub>O emissions corresponds to an additional source of 0.33 Pg CO<sub>2</sub> equivalents (equiv.) yr<sup>−1</sup> from agricultural ecosystems (1 Pg = 10<sup>15</sup> g), and of 0.24 Pg CO<sub>2</sub> equiv. yr<sup>−1</sup> for all other upland ecosystems (Fig. 2). The CO<sub>2</sub>-stimulation of CH<sub>4</sub> emissions corresponds to an additional source of 0.25 Pg CO<sub>2</sub> equiv. yr<sup>−1</sup> from rice paddies and of 0.31 Pg CO<sub>2</sub> equiv. yr<sup>−1</sup> from natural wetlands. Our data indicate a small and non-significant effect of CO<sub>2</sub> on global CH<sub>4</sub> fluxes from upland soils for



**Figure 1** | Results of a meta-analysis of the response of GHG emissions and their potential drivers to rising levels of atmospheric CO<sub>2</sub>. **a**, The effect of increased CO<sub>2</sub> on emissions of N<sub>2</sub>O from upland soil and CH<sub>4</sub> from rice paddies and wetlands. Results are based on 73, 21 and 24 observations, respectively. **b**, The effect of increased CO<sub>2</sub> on root biomass and soil water content. Results are based on 83 and 55 observations, respectively. Effect sizes in all meta-analyses were weighted by replication. Error bars, 95% confidence intervals.

<sup>1</sup>Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona 86011, USA. <sup>2</sup>Merriam-Powell Center for Environmental Research, Northern Arizona University, Flagstaff, Arizona 86011, USA. <sup>3</sup>Department of Botany, School of Natural Sciences, Trinity College Dublin, Dublin 2, Ireland. <sup>4</sup>Department of Biology, University of Florida, Gainesville, Florida 32611-8525, USA.





**Figure 2 | The effect of rising atmospheric CO<sub>2</sub> on GHG emissions, expressed on the global scale.** For N<sub>2</sub>O fluxes, the results for natural and agricultural soils were based on 35 and 19 observations, respectively. For CH<sub>4</sub> fluxes, the results for natural wetlands, rice paddies, natural upland soils and agricultural upland soils were based on 16, 21, 10 and 8 observations, respectively. Effect sizes in all meta-analyses were weighted by replication. Error bars, 95% confidence intervals.

agricultural ecosystems (0.003 Pg CO<sub>2</sub> equiv. yr<sup>-1</sup>) and for all other upland ecosystems (−0.011 Pg CO<sub>2</sub> equiv. yr<sup>-1</sup>). The combined effect of increased CO<sub>2</sub> on emissions of these GHGs is 1.12 Pg CO<sub>2</sub> equiv. yr<sup>-1</sup>.

Rising atmospheric CO<sub>2</sub> is expected to increase soil C storage in terrestrial ecosystems, which may contribute to the current residual C sink on land<sup>7</sup>. Meta-analysis of CO<sub>2</sub> enrichment experiments indicates that the sink is larger for ecosystems receiving fertilizer<sup>12</sup>. Scaled up by the total area of agricultural and non-fertilized ecosystems, these meta-analyses suggest that increased atmospheric CO<sub>2</sub> levels may increase the soil C sink by as much as 4.0 Pg CO<sub>2</sub> yr<sup>-1</sup>. Results presented here indicate that enhanced GHG emissions under increased CO<sub>2</sub> reduce the C mitigation effect of soil C storage by 28% (1.12 Pg/4.0 Pg). The magnitude and significance of this result is insensitive to the choice of the weighting function used in the meta-analysis (Supplementary Fig. 1, Supplementary Table 10).

Experiments included in our database increased atmospheric CO<sub>2</sub> concentration to 630 p.p.m.v. on average, a level expected for the second half of this century<sup>13</sup>. Biogeochemical models predict that at that time, the terrestrial C sink may be as much as 6.8 Pg CO<sub>2</sub> yr<sup>-1</sup> stronger than it is today<sup>4</sup> (when considering forcing by rising CO<sub>2</sub> alone). On the basis of our analysis, a CO<sub>2</sub>-induced rise in GHG fluxes could negate 16.6% (1.12 Pg/6.8 Pg) of the expected increase of the entire terrestrial C sink (Supplementary Table 10).

This estimate (16.6%) is likely to be an underestimate for three reasons. First, most of the studies in our data set measured GHG fluxes during the growing season only, but we assumed these applied to the entire year. Winter emissions of CH<sub>4</sub> in wetlands and rice paddies are typically small<sup>9</sup>; however, winter emissions of N<sub>2</sub>O during freeze-thaw cycles can contribute substantially to annual N<sub>2</sub>O fluxes<sup>14</sup>, and available data indicate that winter emissions of N<sub>2</sub>O are stimulated under increased CO<sub>2</sub> (ref. 15). A recently published data set<sup>16</sup> suggests that N<sub>2</sub>O emissions outside the growing season amount to 88% and 64% of the emissions during the growing season in agricultural systems and natural ecosystems, respectively (see Methods). Assuming that increased CO<sub>2</sub> affects N<sub>2</sub>O emissions proportionately throughout the year, its effect on N<sub>2</sub>O emissions outside the growing season would therefore amount to 0.29 Pg CO<sub>2</sub> equiv. yr<sup>-1</sup> from agricultural systems and 0.15 Pg CO<sub>2</sub> equiv. yr<sup>-1</sup> from natural ecosystems. Together, these fluxes negate an additional 7% of the expected increase of the terrestrial C sink.

Second, atmospheric N deposition is predicted to increase during this century<sup>17</sup>. Because average CO<sub>2</sub> responses of N<sub>2</sub>O emissions were higher in studies receiving additional N (Supplementary Tables 2 and

6), the positive effect of CO<sub>2</sub> on N<sub>2</sub>O emissions may strengthen as ecosystems become enriched in N.

Last, CO<sub>2</sub> effects on N<sub>2</sub>O emissions showed a weak but significant correlation with experiment duration (Supplementary Fig. 2), suggesting that CO<sub>2</sub> effects on N<sub>2</sub>O emissions may increase over time.

Why do GHG emissions respond positively to rising levels of atmospheric CO<sub>2</sub>? Atmospheric CO<sub>2</sub> enrichment increased soil water contents for the studies contributing to our N<sub>2</sub>O database (Fig. 1b, Supplementary Table 11); this result is probably due to improved efficiency of water use by plants, which reduces soil water loss through transpiration<sup>18</sup>. Moreover, increased CO<sub>2</sub> has been shown to enhance soil biological activity across a broad range of ecosystems<sup>12</sup>. Both responses promote soil anoxia, and thus stimulate denitrification<sup>19</sup> (anaerobic microbial respiration of nitrate), one of the major sources of N<sub>2</sub>O from soils<sup>3</sup>. Increased CO<sub>2</sub> also enhanced root biomass in all three habitats (Fig. 1b, Supplementary Table 12). As denitrification is generally stimulated by high availability of labile C as a source of energy<sup>20</sup>, and because new C enters mineral soil mainly through the root system, this increase in root biomass would stimulate denitrification rates—and N<sub>2</sub>O emissions—even further.

Methane is produced only under anaerobic conditions, which are common in soils of rice paddies and natural wetlands but not uplands. Because methanogenic archaea rely on C assimilation by plants as their ultimate source of organic substrates<sup>9</sup>, increased rates of soil C input with increased CO<sub>2</sub> can also stimulate CH<sub>4</sub> emissions. Indeed, the positive correlation between CH<sub>4</sub> emission rates and net ecosystem production in wetlands<sup>21</sup> suggests that plant productivity is a key process in the regulation of CH<sub>4</sub> emission from these ecosystems. The response to increased CO<sub>2</sub> of CH<sub>4</sub> emissions from rice paddies and wetlands showed significant correlation with the CO<sub>2</sub> response of root biomass ( $r^2 = 0.17$ ,  $P = 0.02$ , Supplementary Fig. 6); this further suggests that increased CO<sub>2</sub> stimulates CH<sub>4</sub> production through its positive effect on plant growth and soil C input.

Global changes in climate and atmospheric composition have previously been suggested to affect GHG emissions from natural ecosystems. For instance, a global rise in temperature of 3.4 °C has been predicted to increase CH<sub>4</sub> emissions from wetlands by 78% (ref. 22). In addition to its direct effect on the global climate through radiative forcing, our results identify two indirect mechanisms through which rising atmospheric CO<sub>2</sub> amplifies climate change: by stimulating the release of N<sub>2</sub>O from terrestrial ecosystems, and by enhancing CH<sub>4</sub> release from wetlands and rice paddies. The meta-analytic approach used here, synthesizing results across 49 studies, shows that increased N<sub>2</sub>O and CH<sub>4</sub> emissions are both general and quantitatively important. Future assessments of terrestrial feedbacks to climate change should therefore consider these indirect effects of increased atmospheric CO<sub>2</sub> on the production by soil of trace gases like N<sub>2</sub>O and CH<sub>4</sub>.

## METHODS SUMMARY

We extracted results for soil fluxes of CH<sub>4</sub> and N<sub>2</sub>O, root biomass and soil water contents from CO<sub>2</sub> enrichment studies that were conducted in the field, in growth chambers or in glass houses. Soil fluxes of CH<sub>4</sub> from wetlands, rice paddies and upland soils were considered separately. We divided studies into two categories of N availability based on fertilization rates, that is, more or less than 30 kg N ha<sup>-1</sup> yr<sup>-1</sup>. This cut-off point corresponds to maximum atmospheric N deposition in the United States and most of the European Union<sup>23</sup>. We also made a distinction between studies in pots and field studies, and between studies with planted or natural vegetation. Agricultural ecosystems were defined as cropland and managed grasslands receiving between 30 and 300 kg N ha<sup>-1</sup> yr<sup>-1</sup>.

We quantified the effect of increased CO<sub>2</sub> on GHG fluxes by calculating the natural log of the response ratio ( $R$ ), a metric commonly used in meta-analysis<sup>24</sup>:

$$\ln R = \ln(\text{GHG}_i/\text{GHG}_a)$$

where GHG is the flux of either CH<sub>4</sub> or N<sub>2</sub>O under increased (i) or ambient (a) conditions. We also used  $\ln R$  to assess CO<sub>2</sub> responses of root biomass and soil water contents. We performed our analysis on effect sizes weighted by replication<sup>25</sup>, on

unweighted effect sizes<sup>12</sup>, and on effect sizes weighted by the inverse of the pooled variance<sup>26</sup>.

Treatment effects were also expressed as the difference in annual GHG fluxes on an areal basis ( $U$ ). This metric was essential for upland  $\text{CH}_4$  flux, where values can be both positive and negative (making  $\ln R$  problematic).

We used METAWIN 2.1<sup>27</sup> to generate mean effect sizes and 95% bootstrapped confidence intervals (95% CI). Treatment effects were considered significant if the 95% CI did not overlap with 0. To scale up our results, we multiplied  $U$  by the total vegetated land area covered by each category of experiment<sup>28,29</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 2 September 2010; accepted 9 May 2011.

- Zak, D. R., Pregitzer, K. S., King, J. S. & Holmes, W. E. Elevated atmospheric  $\text{CO}_2$ , fine roots and the response of soil microorganisms: a review and hypothesis. *New Phytol.* **147**, 201–222 (2000).
- Pendall, E. *et al.* Below-ground process responses to elevated  $\text{CO}_2$  and temperature: a discussion of observations, measurement methods, and models. *New Phytol.* **162**, 311–322 (2004).
- Smith, K. A. *et al.* Exchange of greenhouse gases between soil and atmosphere: interactions of soil physical factors and biological processes. *Eur. J. Soil Sci.* **54**, 779–791 (2003).
- Thornton, P. E., Lamarque, J.-F., Rosenbloom, N. A. & Mahowald, N. M. Influence of carbon-nitrogen cycle coupling on land model response to  $\text{CO}_2$  fertilization and climate variability. *Glob. Biogeochem. Cycles* **21**, GB4018, doi:10.1029/2006GB002868 (2007).
- Forster, P. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 129–234 (Cambridge Univ. Press, 2007).
- Long, S. P., Ainsworth, E. A., Rogers, A. & Ort, D. R. Rising atmospheric carbon dioxide: plants FACE the future. *Annu. Rev. Plant Biol.* **55**, 591–628 (2004).
- Gifford, R. M. The global carbon cycle: a viewpoint on the missing sink. *Aust. J. Plant Physiol.* **21**, 1–15 (1994).
- Denman, K. L. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 499–587 (Cambridge Univ. Press, 2007).
- Le Mer, J. & Roger, P. Production, oxidation, emission and consumption of methane by soils: a review. *Eur. J. Soil Biol.* **37**, 25–50 (2001).
- Schulze, E. D. *et al.* Importance of methane and nitrous oxide for Europe's terrestrial greenhouse-gas balance. *Nature Geosci.* **2**, 842–850 (2009).
- Osenberg, C. W., Sarnelle, O., Cooper, S. D. & Holt, R. D. Resolving ecological questions through meta-analysis: goals, metrics, and models. *Ecology* **80**, 1105–1117 (1999).
- Hungate, B. A. *et al.* Assessing the effect of elevated carbon dioxide on soil carbon: a comparison of four meta-analyses. *Glob. Change Biol.* **15**, 2020–2034 (2009).
- Meehl, G. A. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 747–845 (Cambridge Univ. Press, 2007).
- Matzner, E. & Borken, W. Do freeze-thaw events enhance C and N losses from soils of different ecosystems? A review. *Eur. J. Soil Sci.* **59**, 274–284 (2008).
- Kammann, C., Müller, C., Grünhage, L. & Jäger, H.-J. Elevated  $\text{CO}_2$  stimulates  $\text{N}_2\text{O}$  emissions in permanent grassland. *Soil Biol. Biochem.* **40**, 2194–2205 (2008).
- Stehfest, E. & Bouwman, L.  $\text{N}_2\text{O}$  and NO emission from agricultural fields and soils under natural vegetation: summarizing available measurement data and modelling of global annual emissions. *Nutr. Cycl. Agroecosyst.* **74**, 207–228 (2006).
- Galloway, J. N. *et al.* Transformation of the nitrogen cycle: recent trends, questions, and potential solutions. *Science* **320**, 889–892 (2008).
- Wullschlegel, S. D., Tschaplinski, T. J. & Norby, R. J. Plant water relations at elevated  $\text{CO}_2$  – implications for water-limited environments. *Plant Cell Environ.* **25**, 319–331 (2002).
- Smith, M. S. & Tiedje, J. M. Phases of denitrification following oxygen depletion in soil. *Soil Biol. Biochem.* **11**, 261–267 (1979).
- Weier, K. L., Doran, J. W., Power, J. F. & Walters, D. T. Denitrification and the dinitrogen/nitrous oxide ratio as affected by soil water, available carbon, and nitrate. *Soil Sci. Soc. Am. J.* **57**, 66–72 (1993).
- Whiting, G. J. & Chanton, J. P. Primary production control of methane emission from wetlands. *Nature* **364**, 794–795 (1993).
- Shindell, D. T., Walter, B. P. & Faluvegi, G. Impacts of climate change on methane emissions from wetlands. *Geophys. Res. Lett.* **31**, L21202, doi:10.1029/2004GL021009 (2004).
- Holland, E. A., Braswell, B. H., Sulzman, J. & Lamarque, J.-F. Nitrogen deposition onto the United States and western Europe: synthesis of observations and models. *Ecol. Appl.* **15**, 38–57 (2005).
- Hedges, L. V., Gurevitch, J. & Curtis, P. S. The meta-analysis of response ratios in experimental ecology. *Ecology* **80**, 1150–1156 (1999).
- Adams, D. C., Gurevitch, J. & Rosenberg, M. S. Resampling tests for meta-analysis of ecological data. *Ecology* **78**, 1277–1283 (1997).
- Hedges, L. V. & Olkin, I. *Statistical Methods for Meta-Analysis* (Academic, 1985).
- Rosenberg, M. S., Adams, D. C. & Gurevitch, J. METAWIN, Statistical Software for Meta-Analysis Version 2 (Sinauer, 2000).
- World Resources Institute. Land area classification by ecosystem type. (<http://earthtrends.wri.org/datatables/index.php?theme=9>) (2003).
- Aselmann, I. & Crutzen, P. J. Global distribution of natural freshwater wetlands and rice paddies, their net primary productivity, seasonality and possible methane emissions. *J. Atmos. Chem.* **8**, 307–358 (1989).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. A. Prior, G. B. Runion, F. Hagedorn, A. Niboyet, J. C. Blankinship, W. Cheng, T. Kanerva, R. S. Nowak, S. F. Zitzer, F. A. Dijkstra and J. P. Megonigal for sharing their data. Financial support for this study was provided by DOE-NICCR, NSF (DEB-0949460) and the Irish Research Council for Science, Engineering and Technology, co-funded by Marie Curie Actions under FP7.

**Author Contributions** K.J.v.G. and B.A.H. designed the investigation. K.J.v.G. extracted the data from the literature and constructed the database. K.J.v.G. and C.W.O. performed the statistical analyses. All authors contributed to writing the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to K.J.v.G. ([cjvangroenigen@nau.edu](mailto:cjvangroenigen@nau.edu)).

## METHODS

**Data collection.** We extracted results for soil fluxes of CH<sub>4</sub> and N<sub>2</sub>O, root biomass and soil water contents from atmospheric CO<sub>2</sub> enrichment studies, conducted in the field, in growth chambers or in glass houses. We used Google Scholar (Google Inc.) for an exhaustive search of journal articles published before January 2011, using as search terms either “elevated CO<sub>2</sub>” or “CO<sub>2</sub> enrichment”, and either “N<sub>2</sub>O” and “soil”, or “CH<sub>4</sub>”. Further papers were added from a comparable search using Web of Science. For a study to be included in our data set, the atmospheric CO<sub>2</sub> concentration for the ambient and elevated treatments had to be in the range 350–450 p.p.m.v. and 450–800 p.p.m.v., respectively. Means and sample sizes had to be reported for both ambient and elevated CO<sub>2</sub> treatments.

For each study, we noted experimental duration, plant species, N fertilization rates and the type of experimental facility. Estimates of standard deviation were tabulated when available, but were not required for inclusion in the analysis. We included studies involving experiments in pots (that is, any container with dimensions <1 m) or in the field, and studies on natural or planted vegetation. We only considered studies in which soil under both CO<sub>2</sub> treatments had the same treatment history. One study was discarded for this reason. Studies on soil water content and root biomass were only included if data on N<sub>2</sub>O or CH<sub>4</sub> fluxes were available from the same site. When root biomass and soil water content were reported for multiple soil depths, we calculated the overall treatment effects across the entire soil profile. We included separate observations of increased CO<sub>2</sub> effects from a single ecosystem under different experimental treatments (that is, in multifactorial studies). Because wetlands are mostly anaerobic and therefore produce CH<sub>4</sub>, whereas upland soils are mostly aerobic and oxidize CH<sub>4</sub>, these two groups of ecosystems were considered in separate data sets. We also distinguished studies conducted in rice paddies, which like wetlands produce CH<sub>4</sub>. Because the low number of studies on N<sub>2</sub>O fluxes from rice paddies (1) and wetlands (3) did not warrant the construction of separate data sets, these studies were not included in our analysis.

We divided the studies into two categories of N availability based on N fertilization rates, that is, more or less than 30 kg N ha<sup>-1</sup> yr<sup>-1</sup>. This cut-off point was chosen because it is comparable to maximum atmospheric N depositions in the US and most of the EU<sup>23</sup>. We also distinguished between studies on natural or planted vegetation. Agricultural ecosystems were defined as grassland and cropland that received between 30 and 300 kg N ha<sup>-1</sup> yr<sup>-1</sup>. The upper cut-off point was based on reported average fertilization rates for croplands in the world's most intensively fertilized region (that is, East Asia, at 150 kg N ha<sup>-1</sup> yr<sup>-1</sup>)<sup>16</sup>, and the assumption that average fertilizer N use per hectare will be twofold higher in 2050<sup>30</sup>.

**Response metrics.** We evaluated our data sets by using meta-analysis. As a metric for the response of GHG emissions to increased CO<sub>2</sub>, we used the natural log of the response ratio<sup>24</sup>. This metric starts with an estimate of the relative change in GHG emissions between ambient and increased CO<sub>2</sub> treatments, and log-transforms it to improve its statistical behaviour.

$$\ln R = \ln(\text{GHG}_i/\text{GHG}_a)$$

where GHG is the flux of either CH<sub>4</sub> or N<sub>2</sub>O under increased (i) or ambient (a) conditions. We also used lnR as a metric for CO<sub>2</sub> responses of root biomass and soil water contents. Fluxes of CH<sub>4</sub> from upland soils could not be analysed using this metric, because our data set included both sites with negative (that is, CH<sub>4</sub> uptake) and positive (CH<sub>4</sub> emissions) fluxes. For this reason, we also used the difference in annual emissions, expressed on an areal basis (*U*) as a metric:

$$U = (\text{GHG}_i - \text{GHG}_a)$$

with GHG<sub>i</sub> and GHG<sub>a</sub> as before. All but one study on wetland soils found net CH<sub>4</sub> emissions under both ambient and increased CO<sub>2</sub> conditions (Supplementary Data 2). This one study, which reported that increased CO<sub>2</sub> turned wetland soils from a net sink of CH<sub>4</sub> into a net source, was therefore excluded when calculating lnR, but included when calculating *U*.

Several studies only measured N<sub>2</sub>O and CH<sub>4</sub> fluxes during the growing season. In these cases, we assumed that the effect of increased CO<sub>2</sub> on annual fluxes occurred entirely during this period. When the length of the growing season was not explicitly indicated, we assumed a growing season of 150 days. When studies measured gas fluxes for multiple years, fluxes were averaged over time.

**Weighting functions.** We performed analyses using non-parametric weighting functions and generated confidence intervals (CIs) on weighted effects sizes using bootstrapping. Because effect size estimates and subsequent inferences in meta-analysis may depend on how individual studies are weighted<sup>12</sup>, we used three different weighting functions. First, weighted by replication:  $W_R = (n_a \times n_i)/(n_a + n_i)$ , where  $n_a$  and  $n_i$  are the number of replicates under ambient and increased CO<sub>2</sub>, respectively<sup>25</sup>. For pot studies, *n* equalled the number of replicate experimental

facilities (that is, growth chambers, glass houses, and so on), rather than the number of pots per CO<sub>2</sub> treatment. Second, unweighted. Each observation was assigned an equal weight:  $W_U = 1$ . Third, weighted by the inverse of the pooled variance, the weighting function conventionally used in meta-analyses<sup>26</sup>:  $W_V = 1/(\text{var}_a/\text{GHG}_a^2 + \text{var}_i/\text{GHG}_i^2)$ , with GHG<sub>a</sub> and GHG<sub>i</sub> as before, and var<sub>a</sub> and var<sub>i</sub> as their respective variance.

When variance estimates were missing for a study, we calculated the average coefficient of variation (CV) within each data set, and then approximated the missing variance by multiplying the reported mean by the average CV and squaring the result.

When multiple effects were extracted from the same experimental site, we adjusted the weights defined above by the total number of observations from that site. This approach ensured that all experimental comparisons in multifactor studies could be included in the data set without dominating the overall effect size. For three experimental sites, multiple studies were done on the same GHG fluxes at different points in time. We adjusted the weights of observations from these studies by the total number of observations per site. Thus, the final weights used in the analyses were  $w_{f,i} = W_{f,i}/n_c$  where  $n_c$  was the number of observations from the same site as the *i*th observation, and *f* was the index that referred to one of the three weighting functions defined above.

Mean effects sizes ( $\ln \bar{R}$ ,  $\bar{U}$ ) for different categories of studies were estimated as:

$$\ln \bar{R} = \frac{\sum_i (\ln R_i \times w_{f,i})}{\sum_i w_{f,i}}$$

$$\bar{U} = \frac{\sum_i (U_i \times w_{f,i})}{\sum_i w_{f,i}}$$

We used METAWIN 2.1<sup>27</sup> to generate these mean effect sizes and 95% bootstrapped CIs (4,999 iterations). Treatment effects were considered significant if the 95% CI did not overlap with 0. The results for the analyses on lnR were back-transformed and reported as percentage change under increased CO<sub>2</sub> (that is,  $100 \times (R - 1)$ ) to ease interpretation.

We tested whether lnR for GHG emissions was correlated with lnR for root biomass using the statistical package SPSS 19. Similarly, we tested whether lnR for GHG emissions was correlated with experiment duration or the level of CO<sub>2</sub> enrichment. The effect of increased CO<sub>2</sub> on soil emissions of N<sub>2</sub>O, but not CH<sub>4</sub>, showed a weak positive correlation with experiment duration (Supplementary Figs 2 and 3). lnR was not significantly correlated with the degree of CO<sub>2</sub> enrichment for either N<sub>2</sub>O or CH<sub>4</sub> emissions (Supplementary Figs 4 and 5). This result is probably due to the large variation in treatment effects between studies, masking effects of the degree in CO<sub>2</sub> enrichment. Alternatively, the results may reflect that plant growth is a saturating function of CO<sub>2</sub> concentrations. Since experiments increased atmospheric CO<sub>2</sub> to a similar extent for all data sets (Supplementary Table 13), we did not normalize effect sizes for the level of CO<sub>2</sub> enrichment.

Results using the different weighting functions were qualitatively similar. However, the variance-based weighting function,  $W_V$ , yielded weights that varied over 1,000 times in magnitude (Supplementary Data 1 and 2). By assigning extreme importance to individual observations, average effect sizes were largely determined by a small number of studies. Because variance estimates are notoriously unreliable (especially given the small samples common in many of these studies), we favoured the use of the alternative weighting functions (which assigned less extreme weights). In this Letter, we provide results of the analyses on effect sizes that were weighted by replication; results for all weighting functions can be found in Supplementary Tables 2–8, 11 and 12.

**Scaling of results.** We scaled up the results from the experiments by multiplying them by the total land area covered by the particular type of habitat that was being summarized. In other words, we took the mean effects and confidence intervals for *U* calculated above and scaled them:

$$F = \bar{U} \times H$$

where *F* is expressed in Pg CO<sub>2</sub> equiv. yr<sup>-1</sup>, and *H* is the amount of habitat in uplands, wetlands, or rice paddies (103.1, 5.7, and 1.3 million km<sup>2</sup>, respectively<sup>28,29</sup>). Because N fertilization increases N<sub>2</sub>O emissions<sup>16,17</sup> and enhances plant growth, we distinguished between upland agricultural ecosystems (that is, 19.0 million km<sup>2</sup> of fertilized grasslands and croplands<sup>16</sup>, minus 1.3 million km<sup>2</sup> of rice paddies<sup>28</sup>) and ecosystems receiving little or no fertilizer (103.1 – 19.0 + 1.3 = 85.4 million km<sup>2</sup>).

We estimated the contribution of winter N<sub>2</sub>O emissions to total N<sub>2</sub>O emissions from a recently published data set<sup>16</sup>. For agricultural soils and soils under natural vegetation, studies conducted over the growing season and lasting 100–200 days were compared to studies conducted over the entire year (that is, lasting >300 days). Because tropical and subtropical systems do not experience marked



growing seasons, we excluded studies from those regions. For agricultural soils, we only considered studies on grassland and cropland receiving 30–300 kg N ha<sup>-1</sup> yr<sup>-1</sup> (that is, the same restrictions that applied to our data sets 1 and 2 for the global extrapolation shown in Fig. 2). The difference in mean N<sub>2</sub>O emissions between the two categories of study duration was assumed to be representative of N<sub>2</sub>O emissions outside the growing season.

To estimate the CI for the combined effect of increased CO<sub>2</sub> on all six GHG fluxes shown in Fig. 2, we calculated the square root of the sum of the squared CIs. Because the original CIs were asymmetric, we did this separately for the upper and lower CIs. All studies on rice paddies were conducted on planted vegetation, experimental conditions resembling real-world conditions. When we combined

our extrapolated data to calculate the overall CO<sub>2</sub> effect on CH<sub>4</sub> emissions, we therefore included all available data from rice paddies (Fig. 2, Supplementary Fig. 1). To compare the emissions of GHG with soil C sequestration under increased CO<sub>2</sub>, we used results from the analyses weighted by replication and from unweighted analyses as reported in ref. 12, applying the same study selection criteria as for studies in our current data set. These results were expressed as a function of total land area, using the same approach that was used to scale up our results on GHG fluxes.

30. Tilman, D. *et al.* Forecasting agriculturally driven global environmental change. *Science* **292**, 281–284 (2001).

# In vivo genome editing restores haemostasis in a mouse model of haemophilia

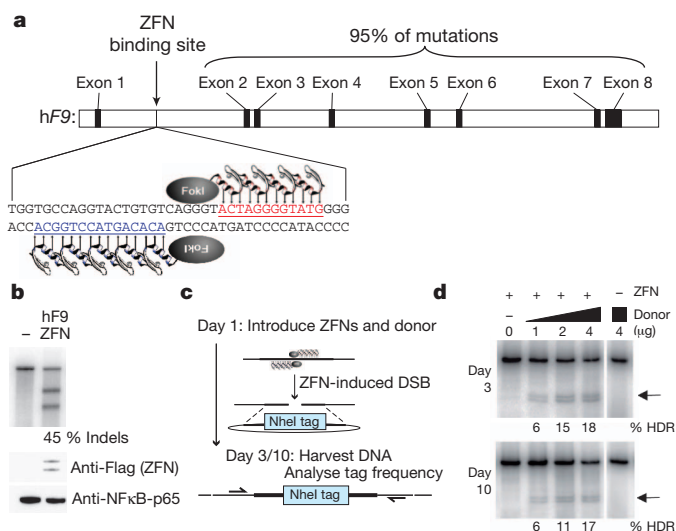
Hojun Li<sup>1</sup>, Virginia Haurigot<sup>1</sup>, Yannick Doyon<sup>2</sup>, Tianjian Li<sup>2</sup>, Sunnie Y. Wong<sup>2</sup>, Anand S. Bhagwat<sup>1</sup>, Nirav Malani<sup>3</sup>, Xavier M. Anguela<sup>1</sup>, Rajiv Sharma<sup>1</sup>, Lacramiora Ivanciu<sup>1</sup>, Samuel L. Murphy<sup>1</sup>, Jonathan D. Finn<sup>1</sup>, Fayaz R. Khazi<sup>1</sup>, Shangzhen Zhou<sup>1</sup>, David E. Paschon<sup>2</sup>, Edward J. Rebar<sup>2</sup>, Frederic D. Bushman<sup>3</sup>, Philip D. Gregory<sup>2</sup>, Michael C. Holmes<sup>2</sup> & Katherine A. High<sup>1,4</sup>

Editing of the human genome to correct disease-causing mutations is a promising approach for the treatment of genetic disorders. Genome editing improves on simple gene-replacement strategies by effecting *in situ* correction of a mutant gene, thus restoring normal gene function under the control of endogenous regulatory elements and reducing risks associated with random insertion into the genome. Gene-specific targeting has historically been limited to mouse embryonic stem cells. The development of zinc finger nucleases (ZFNs) has permitted efficient genome editing in transformed and primary cells that were previously thought to be intractable to such genetic manipulation<sup>1</sup>. *In vitro*, ZFNs have been shown to promote efficient genome editing via homology-directed repair by inducing a site-specific double-strand break (DSB) at a target locus<sup>2–4</sup>, but it is unclear whether ZFNs can induce DSBs and stimulate genome editing at a clinically meaningful level *in vivo*. Here we show that ZFNs are able to induce DSBs efficiently when delivered directly to mouse liver and that, when co-delivered with an appropriately designed gene-targeting vector, they can stimulate gene replacement through both homology-directed and homology-independent targeted gene insertion at the ZFN-specified locus. The level of gene targeting achieved was sufficient to correct the prolonged clotting times in a mouse model of haemophilia B, and remained persistent after induced liver regeneration. Thus, ZFN-driven gene correction can be achieved *in vivo*, raising the possibility of genome editing as a viable strategy for the treatment of genetic disease.

Viral-vector-mediated transfer of the wild-type copy of a gene that is defective in disease (gene replacement therapy) has been performed successfully in a variety of animal models and in humans<sup>5–9</sup>. However, disadvantages of gene replacement include risks related to insertional mutagenesis<sup>10–12</sup> and loss of endogenous regulatory signals that control gene expression. Gene-specific targeting in mouse induced pluripotent stem cells has highlighted the potential to overcome these challenges through *ex vivo* correction of a disease-causing mutation<sup>13</sup>. However, most genetic diseases affect organ systems in which *ex vivo* manipulation of target cells is not feasible. One such organ is the liver, the major site of synthesis of plasma proteins, including blood coagulation factors. A model genetic disease for gene therapy in the liver is haemophilia B, which is caused by deficiency of blood coagulation factor IX, encoded by the *F9* gene. Most affected individuals have circulating levels of factor IX that are below 1% of normal (5,000 ng ml<sup>-1</sup>), but restoration to about 5% activity (250 ng ml<sup>-1</sup>) converts severe haemophilia B to a mild form<sup>14</sup>. Most mutations in the *F9* gene are distributed across the coding sequences of exons 2–8 (Fig. 1a)<sup>15</sup>. Thus, specific targeting of any single mutant allele would not allow complete coverage of the wide spectrum of mutations found in the human population. However, ZFN-mediated targeting of a promoterless therapeutic gene fragment<sup>2,16</sup> (that is, a partial cDNA preceded by a splice acceptor site) into the first intron

of *F9* would allow for splicing of a wild-type coding sequence with exon 1, leading to expression of functionally active factor IX and rescue of the defect caused by most mutations. We therefore sought to investigate whether ZFNs combined with a targeting vector carrying the wild-type *F9* exons 2–8 could induce gene targeting *in vivo* and correct a mutated *F9* gene *in situ*.

We designed ZFNs targeting intron 1 of the human *F9* (hF9) gene (F9 ZFNs, Supplementary Fig. 1) and confirmed their capacity to introduce a DSB at the intended target site (Fig. 1b) and to stimulate genome editing by homology-directed repair (HDR) in human erythroleukaemia K-562 cells (Fig. 1c, d). This ZFN pair was highly



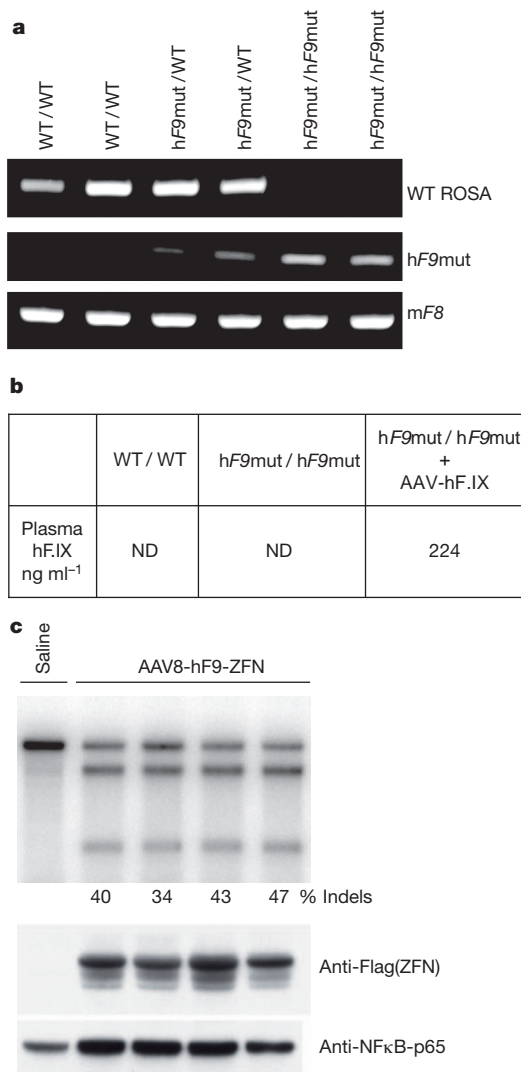
**Figure 1 | F9 ZFNs cleave human *F9* intron 1 and induce homology-directed repair *in vitro*.** **a**, F9 ZFNs target intron 1 of the human *F9* gene, allowing homology-directed repair upstream of 95% of *F9* mutations. **b**, K-562 cells were transfected with ZFN expression constructs (400 ng, right lane) or not transfected (left lane) and genomic DNA was harvested 3 d after transfection. The Cel-I assay was used to determine the frequency of ZFN-induced indels in both samples, indicated as '% Indels' below the right lane. Expression of Flag-tagged ZFN is confirmed by anti-Flag immunoblotting and anti-NFκB-p65 serves as a loading control. **c**, Schematic of RFLP assay, detailing ZFN-mediated targeting of a *NheI* restriction-site tag to the human *F9* gene. **d**, Co-transfection of 400 ng of ZFN expression plasmid with increasing amounts of *NheI* donor plasmid (0–4 μg) results in increasing levels of HDR at day 3 and day 10 after transfection, whereas transfection of the *NheI* donor alone (4 μg) does not result in detectable HDR. Black arrows denote *NheI*-sensitive cleavage products resulting from HDR. PCR was performed using <sup>32</sup>P-labelled nucleotides, followed by polyacrylamide gel electrophoresis (PAGE) and band-intensity quantification by autoradiography. Lanes with no quantification had no detectable HDR.

<sup>1</sup>Division of Hematology, CTRB 5000, Children's Hospital of Philadelphia, 3501 Civic Center Boulevard, Philadelphia, Pennsylvania 19104, USA. <sup>2</sup>Sangamo BioSciences, Point Richmond Tech Center, 501 Canal Boulevard, Suite A100, Richmond, California 94804, USA. <sup>3</sup>Department of Microbiology, 426 Johnson Pavilion, University of Pennsylvania School of Medicine, 3610 Hamilton Walk, Philadelphia, Pennsylvania 19104, USA. <sup>4</sup>Howard Hughes Medical Institute, 415 Curie Boulevard, Philadelphia, Pennsylvania 19104, USA.

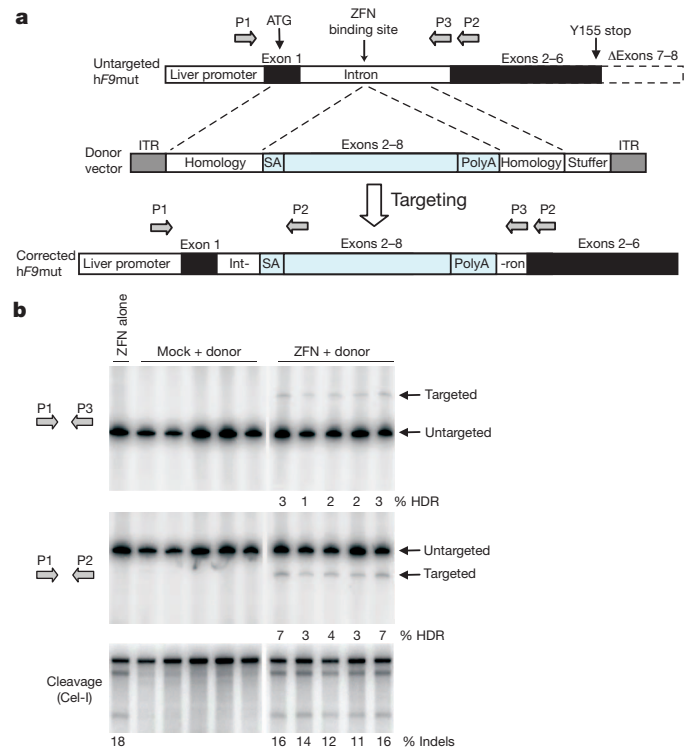
active, driving small insertions and/or deletions (indels), characteristic of DSB repair by non-homologous end-joining (NHEJ), in up to 45% of alleles, and stable integration of the NheI restriction site in ~17–18% of alleles. This latter event is diagnostic of repair by HDR, using a homologous donor template designed to insert a novel restriction enzyme site into the *F9* locus. Similar results were obtained in the Hep3B human hepatocyte line (Supplementary Fig. 2). For *in vivo* evaluation, we generated a humanized mouse model of haemophilia B because the *F9* ZFNs target a site in intron 1 of *hF9* that is absent from the murine gene. We constructed an *hF9* mini-gene<sup>17</sup>, under the

control of a liver-specific enhancer and promoter<sup>18</sup>, that mimics a previously identified mutation (Y155stop)<sup>19</sup>, resulting in the absence of circulating factor IX protein. We knocked in this mini-gene at the mouse *ROSA26* locus<sup>20</sup>, confirmed its genotype (Fig. 2a) and showed that the resulting transgenic mice had no detectable circulating human factor IX (Fig. 2b). We then crossed these mice (hereafter referred to as *hF9mut* mice) with an existing mouse model that has a deletion of the murine *F9* gene<sup>21</sup>, generating *hF9mut*/HB mice to test ZFN-driven gene correction activity *in vivo* (Fig. 3a).

To deliver the *F9* ZFNs to the liver, we generated a hepatotropic adeno-associated virus vector, serotype 8 (AAV8-ZFN) expressing the *F9* ZFNs from a liver-specific enhancer and promoter<sup>18</sup>. To test the cleavage activity of the *F9* ZFNs *in vivo*, we injected *hF9mut* mice through the tail vein with AAV8-ZFN and isolated liver DNA at day 7 after injection. Cleavage activity was measured via the surveyor nuclease



**Figure 2 | AAV8-mediated delivery of *F9* ZFNs to *hF9mut* mouse liver results in cleavage of *hF9mut* intron 1 *in vivo*.** **a**, PCR genotyping of the parental strain (WT), a mouse heterozygous for the *hF9* mutant construct knocked into the *ROSA26* locus (*hF9mut*/WT), and a mouse homozygous for *hF9mut* knocked into the *ROSA26* locus (*hF9mut*/h*F9mut*). The murine factor VIII (mF8) PCR product indicates no inhibition of PCR. **b**, Human factor IX (hF.IX) levels in plasma, assayed by human factor IX ELISA, in wild-type mice, homozygous *hF9mut* mice and *hF9mut* mice injected with a viral vector expressing human factor IX ( $1 \times 10^9$  vector genomes (v.g.) AAV-human factor IX<sup>20</sup>, injected via the tail vein). ND, none detected. **c**, Tail-vein injection of  $1 \times 10^{11}$  v.g. AAV8-ZFN expression vector into *hF9mut* mice results in cleavage of intron 1. The Cel-I assay was performed on liver DNA, isolated at day 7 after injection, to determine the frequency of ZFN-induced indels, indicated as '% Indels' below each lane, resulting from cleavage of the *hF9mut* intron. Lane with no quantification had no detectable cleavage products. Each lane represents an individual mouse. Expression of Flag-tagged ZFN was confirmed by anti-Flag immunoblotting of whole-liver lysate.



**Figure 3 | *F9* ZFNs promote AAV-mediated targeting of wild-type *F9* exons 2–8 to *hF9mut* intron 1 *in vivo*.** **a**, The *hF9mut* gene mutation (truncation of exons 7 and 8) can be bypassed by targeted integration of *hF9* exons 2–8 into intron 1. Targeted and untargeted *hF9mut* alleles can be differentiated by PCR using primers P1, P2 and P3. The locations of the start codon and premature stop mutation are indicated by arrows. The left arm of homology spans from the beginning of exon 1 to the ZFN target site. (Deletion of exon 1 from the left homology arm does not alter results, see Supplementary Fig. 13.) The right arm of homology spans intronic sequence 3' of the ZFN target site. polyA, polyadenylation site; SA, splice acceptor site. **b**, PCR analysis with primer pairs P1/P3 (upper panel) and P1/P2 (middle panel), showing successful gene targeting by HDR after intraperitoneal co-injection of  $5 \times 10^{10}$  v.g. AAV8-ZFN and  $2.5 \times 10^{11}$  v.g. AAV8-donor in *hF9mut*/HB mice at day 2 of life ( $n = 5$ ), but not after injection of  $5 \times 10^{10}$  v.g. AAV8-ZFN alone ( $n = 1$ ) or co-injection of  $5 \times 10^{10}$  v.g. AAV8-mock and  $2.5 \times 10^{11}$  v.g. AAV8-donor ( $n = 5$ ). The mock vector replaces *F9* ZFN coding sequences with renilla luciferase. PCR was performed using <sup>32</sup>P-labelled nucleotides, followed by PAGE and quantification of product-band intensity by autoradiography to evaluate targeting frequency. Targeting frequencies are rounded down to the nearest whole number. Lower panel: intraperitoneal injection of AAV8-ZFN expression vector into *hF9mut* mice results in cleavage of intron 1. The Cel-I assay was performed on liver DNA to determine the frequency of ZFN-induced indels, indicated as '% Indels' below each lane, resulting from cleavage of the *hF9mut* intron. Lanes with no quantification had no detectable HDR or indels. Each lane represents an individual mouse.



(Cel-I) assay<sup>22</sup> which determines the frequency of indels that are characteristic of DSB repair by NHEJ. We observed mutation frequencies ranging from 34% to 47%, demonstrating that coupling of the F9 ZFNs with AAV8-mediated delivery promotes highly efficient genome modification in mouse liver (Fig. 2c). These results were confirmed by direct sequencing of the target locus (Supplementary Fig. 3).

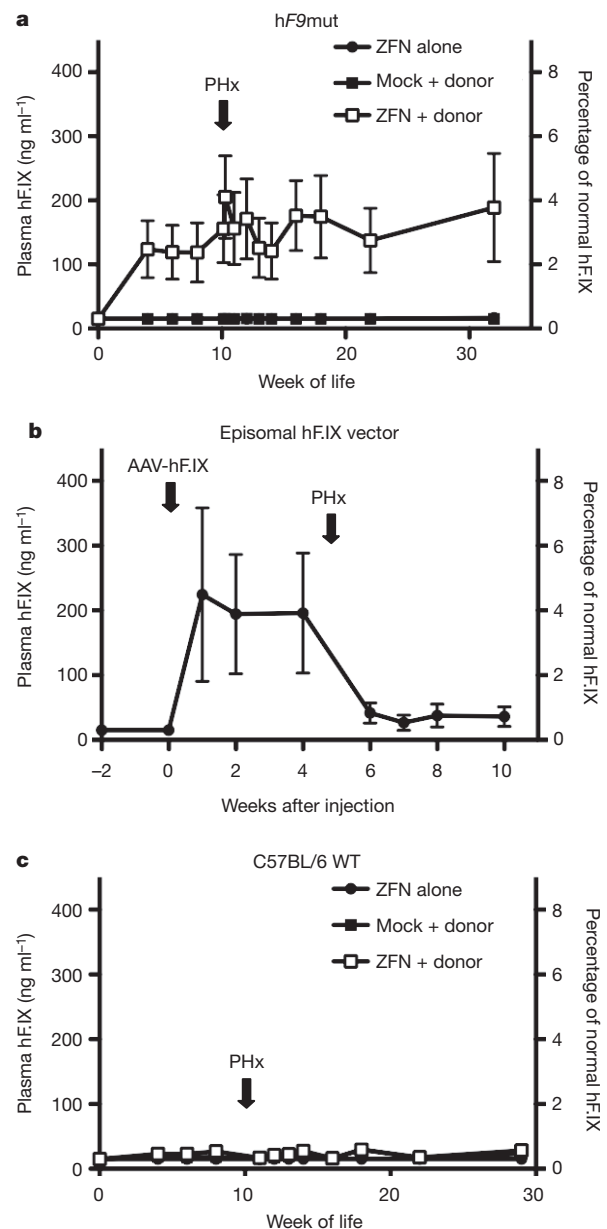
To correct the mutated *hF9* gene *in situ*, we generated an AAV donor template vector (AAV8-donor) for gene targeting, with arms of homology flanking a corrective, partial cDNA cassette containing exons 2–8 of the wild-type *hF9* gene, flanked by splice-acceptor and poly-adenylation sites (Fig. 3a). Having established that we could detect HDR readily *in vitro* (Supplementary Fig. 4), we co-injected *hF9*mut/ HB mice by intraperitoneal injection at day 2 of life with AAV8-ZFN + AAV8-donor, AAV8-mock + AAV8-donor or AAV8-ZFN alone. (Note that intraperitoneal injection in neonatal mice is less efficient than tail-vein injection in adult mice (compare Cel-I results in Fig. 3b to those in Fig. 2c) but it is used because it leads to higher survival rates.) At week 10 of life, we extracted liver DNA to assay gene replacement at the *hF9* locus via HDR. Using primers that hybridize to the chromosome outside the donor homology arms, generating a larger amplicon for a targeted allele (Fig. 3a, primers P1/P3 and Fig. 3b, upper panel), we observed HDR only in mice receiving both the donor and F9 ZFNs, with targeting efficiencies in the 1–3% range (Fig. 3b, upper panel). We confirmed HDR using alternative primers that hybridize to sites outside the donor homology arms and within the inserted cassette, respectively (Fig. 3a, primers P1/P2 and Fig. 3b, middle panel). Thus, co-delivery of ZFNs and a donor template, using AAV vectors, leads to HDR *in vivo*.

To determine whether ZFN-mediated gene targeting results in production of circulating human factor IX, we injected *hF9*mut mice intraperitoneally at day 2 of life with AAV8-ZFN alone, AAV8-mock + AAV8-donor or AAV8-ZFN + AAV8-donor. Human factor IX levels in the plasma of mice receiving ZFN alone or mock + donor averaged  $<15 \text{ ng ml}^{-1}$  (the lower limit of detection of the assay), whereas mice receiving ZFN + donor averaged  $116\text{--}121 \text{ ng ml}^{-1}$ , corresponding to 2–3% of normal levels (Fig. 4a): significantly more than mice receiving ZFN alone and mice receiving mock + donor ( $P \leq 0.006$  at all time points, 2-tailed *t*-test, Supplementary Fig. 5). Notably, in individual mice, the amount of circulating human factor IX correlated directly with the detected level of gene targeting via HDR (Supplementary Fig. 6).

To confirm stable genomic correction, we performed partial hepatectomies. Levels of human factor IX persisted after hepatectomies performed after genome editing (Fig. 4a), whereas an episomal AAV vector expressing human factor IX (AAV-human factor IX, Fig. 4b) showed markedly reduced human factor IX expression after hepatectomy, because extra-chromosomal episomes are lost during liver regeneration<sup>23</sup> (Fig. 4b). Control mice receiving ZFN alone or mock + donor continued to average  $<15 \text{ ng ml}^{-1}$  after hepatectomy (Fig. 4a) ( $P \leq 0.01$  at all time points, 2-tailed *t*-test, Supplementary Fig. 5).

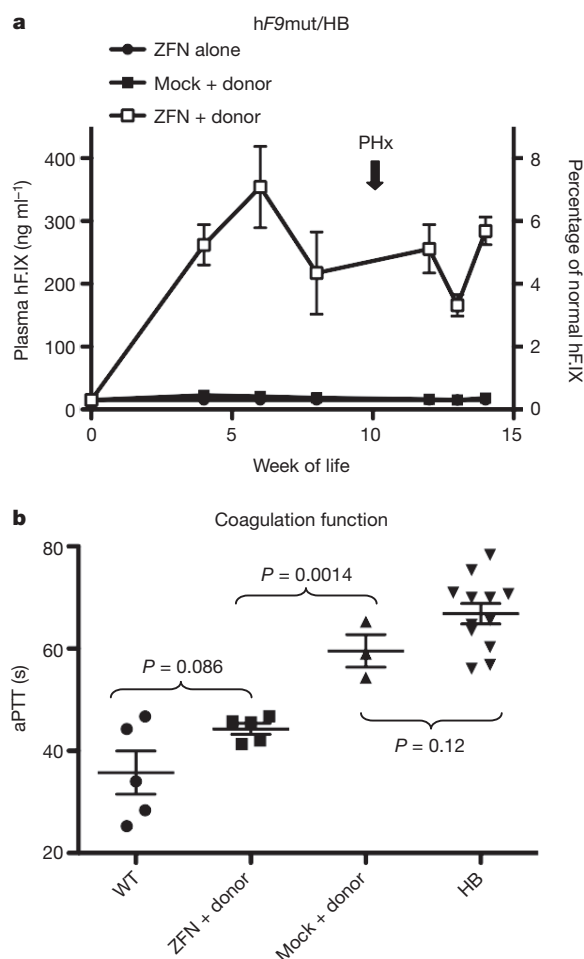
To ensure that the expression of human factor IX did not result from random donor integration into the genome, we injected wild-type mice (lacking the *hF9*mut mini-gene) intraperitoneally at day 2 of life with AAV8-ZFN alone, AAV8-mock + AAV8-donor or AAV8-ZFN + AAV8-donor. Notably, human factor IX levels in the plasma of mice in these groups averaged  $<15 \text{ ng ml}^{-1}$ ,  $<30 \text{ ng ml}^{-1}$  and  $<30 \text{ ng ml}^{-1}$ , respectively (Fig. 4c), indicating that most of the expression of human factor IX in *hF9*mut mice treated with ZFN + donor came from specific gene correction. PCR targeting assays in these wild-type control mice were negative, indicating that amplicons used to quantify HDR were target-gene-specific (Supplementary Fig. 7).

To determine whether ZFN-mediated gene targeting would provide circulating levels of human factor IX that were sufficient to correct the haemophilia B phenotype, we injected *hF9*mut/ HB mice intraperitoneally at day 2 of life with AAV8-ZFN alone, AAV8-mock + AAV8-donor or AAV8-ZFN + AAV8-donor. Levels of human factor IX in



**Figure 4** | *In vivo* *hF9*mut gene correction results in stable circulating factor IX. **a**, Levels of human factor IX in plasma of *hF9*mut mice after intraperitoneal injection at day 2 of life with either  $5 \times 10^{10}$  v.g. AAV8-ZFN alone ( $n = 7$ ),  $5 \times 10^{10}$  v.g. AAV8-ZFN and  $2.5 \times 10^{11}$  v.g. AAV8-donor ( $n = 7$ ), or  $5 \times 10^{10}$  v.g. AAV8-mock and  $2.5 \times 10^{11}$  v.g. AAV8-donor ( $n = 6$ ). Partial hepatectomy (PHx) was performed at the time indicated by the arrow. Plasma levels of human factor IX were assayed by ELISA. Error bars denote s.e.m. **b**, Levels of human factor IX in plasma of wild-type mice ( $n = 3$ ) after tail-vein injection of  $1 \times 10^9$  v.g. AAV-human factor IX (predominantly episomal), with subsequent PHx. Plasma levels of human factor IX were assayed by ELISA. Error bars denote s.e.m. **c**, Levels of human factor IX in plasma of wild-type C57BL/6 mice after intraperitoneal injection at day 2 of life with either  $5 \times 10^{10}$  v.g. AAV8-ZFN alone ( $n = 8$  before PHx,  $n = 4$  after PHx),  $5 \times 10^{10}$  v.g. AAV8-ZFN and  $2.5 \times 10^{11}$  v.g. AAV8-donor ( $n = 9$  before PHx,  $n = 5$  after PHx), or  $5 \times 10^{10}$  v.g. AAV8-mock and  $2.5 \times 10^{11}$  v.g. AAV8-donor ( $n = 6$  before PHx,  $n = 5$  after PHx). Plasma levels of human factor IX were assayed by ELISA. Error bars denote s.e.m.

the plasma of mice receiving ZFN alone again averaged  $<15 \text{ ng ml}^{-1}$ . Mice receiving mock + donor averaged  $<25 \text{ ng ml}^{-1}$  and mice receiving ZFN + donor had significantly higher levels of human factor IX ( $P \leq 0.04$  at all time points compared to mock + donor, 2-tailed *t*-test, Supplementary Fig. 5), averaging  $166\text{--}354 \text{ ng ml}^{-1}$ , 3–7% of normal circulating levels (Fig. 5a). A titration of AAV-donor showed that the



**Figure 5 | Hepatic hF9mut gene correction results in phenotypic correction of haemophilia B.** **a**, Levels of human factor IX in plasma of hF9mut/HB mice after intraperitoneal injection at day 2 of life with either  $5 \times 10^{10}$  v.g. AAV8-ZFN alone ( $n = 10$  before PHx,  $n = 1$  after PHx),  $5 \times 10^{10}$  v.g. AAV8-ZFN and  $2.5 \times 10^{11}$  v.g. AAV8-donor ( $n = 9$  before PHx,  $n = 5$  after PHx), or  $5 \times 10^{10}$  v.g. AAV8-mock and  $2.5 \times 10^{11}$  v.g. AAV8-donor ( $n = 9$  before PHx,  $n = 3$  after PHx). Plasma levels of human factor IX were assayed by ELISA. Error bars denote s.e.m. **b**, Test of clot formation by aPTT at week 14 of life in mice that had received intraperitoneal injection at day 2 of life with  $5 \times 10^{10}$  v.g. AAV8-ZFN and  $2.5 \times 10^{11}$  v.g. AAV8-donor ( $n = 5$ ) or  $5 \times 10^{10}$  v.g. AAV8-mock and  $2.5 \times 10^{11}$  v.g. AAV8-donor ( $n = 3$ ). The aPTTs of wild-type (WT,  $n = 5$ ) and haemophilia B (HB,  $n = 12$ ) mice are shown for comparison. *P*-values are from 2-tailed Student's *t*-test of WT versus ZFN + donor, ZFN + donor versus mock + donor and mock + donor versus HB. Error bars denote s.e.m.

degree of correction was dependent on the dose of AAV-donor (Supplementary Fig. 8). To determine whether the haemophilia B phenotype was corrected, we assayed activated partial thromboplastin time (aPTT), a measure of clot-formation kinetics that is markedly prolonged in haemophilia. The average aPTTs for wild-type mice ( $n = 5$ ) and haemophilia B mice ( $n = 12$ ) were 36 s and 67 s, respectively (Fig. 5b). Mice receiving mock + donor ( $n = 3$ ) averaged 60 s, whereas mice receiving ZFN + donor ( $n = 5$ ) had significantly shortened aPTTs, averaging 44 s ( $P = 0.0014$  compared to mock + donor, 2-tailed *t*-test). Clotting times for ZFN + donor and wild-type mice were not significantly different ( $P = 0.086$ , 2-tailed *t*-test, Fig. 5b). Together, these data demonstrate a clinically significant correction of the coagulation defect in haemophilia B, via direct *in vivo* delivery of ZFNs to mediate permanent correction of the genome in mouse hepatocytes.

To begin to evaluate the specificity of this approach, we used a method based on the systematic evolution of ligands by exponential enrichment (SELEX)<sup>22</sup> to identify the top 20 potential off-target sites for the F9 ZFNs in the mouse genome. Cel-I assays performed at each

of these sites were unable to detect cleavage in 19 out of 20 (lower limit of detection 1%). At the twentieth site, located in an intergenic region at mouse chromosome 9qE3.1, we detected cleavage at a tenth the frequency seen at the F9 target site (Supplementary Fig. 9). Thus, the specificity of the hF9 ZFNs is comparable to CCR5-specific ZFNs, by this analysis<sup>22</sup>.

To investigate the specificity of the ZFN approach further, we used ligation-mediated PCR and 454 pyrosequencing to detect sites of AAV vector integration genome-wide<sup>24</sup>. A comparison of ZFN + donor and mock + donor mice revealed similar distributions of AAV integration sites across the mouse genome (Supplementary Fig. 10); this integration site distribution was consistent with previously reported data showing that genes<sup>24,25</sup>, but not oncogenes, were favoured as integration sites. We next validated the prediction from *in vitro* studies<sup>26</sup> that a ZFN-induced DSB would capture the AAV vector itself, by employing a direct PCR approach using primers that anneal to the hF9mut locus and the AAV inverted terminal repeat (ITR) (Supplementary Fig. 11). This assay confirmed AAV integration at the ZFN target site in ZFN + donor mice but not in mock + donor mice. Finally, a pre-clinical evaluation of toxicity in injected and control mice showed no effects on growth or weight gain in either hF9mut or wild-type mice ( $n = 43$ ) over 8 months of observation (data not shown), and no changes in liver function tests at 4, 29 and 32 weeks after injection (Supplementary Fig. 12), indicating that the treatment was well tolerated.

Studies showing that ZFNs can mediate gene correction efficiently through the introduction of site-specific DSBs, and can induce HDR in cultured cells, have provided important proof-of-concept results for the clinical application of engineered nucleases for diseases affecting cells that can be removed and returned to the patient. However, the necessity to isolate and manipulate cells *ex vivo* limits the application of this technology to a subset of genetic diseases. Our results show that AAV-mediated delivery of a donor template and ZFNs *in vivo* induces gene targeting, resulting in measurable circulating levels of factor IX. This therapeutic strategy is sufficient to restore haemostasis in a mouse model of haemophilia B, thus demonstrating genome editing in an animal model of a disease. Clinical translation of these results will require optimization of correction efficiency and a thorough analysis of off-target effects in the human genome, an issue that we have begun to monitor. Together, these data show that AAV-mediated delivery of ZFNs and a donor template gives rise to persistent and clinically meaningful levels of genome editing *in vivo*, and thus can be an effective strategy for targeted gene disruption or *in situ* correction of genetic disease *in vivo*.

## METHODS SUMMARY

Zinc finger nucleases targeting the hF9 gene were designed and validated as described in Methods. ZFN expression, donor template and AAV-vector-production plasmids were constructed using standard molecular biology techniques. AAV vectors were produced through triple transfection of HEK 293T cells. K-562, Hep3B and HEK 293T cells were cultured and transfected using standard techniques. hF9mut mice were created by targeted transgenesis as described in Methods. Mouse injections, plasma collection and surgical procedures were approved by the Children's Hospital of Philadelphia institutional animal care and use committee, and performed as described in Methods. The Cel-I assay, target-site sequencing, restriction-fragment length polymorphism (RFLP) knock-in assay, targeting assay, human factor IX enzyme-linked immunosorbent assay (ELISA), aPTT, liver function tests, SELEX, ligation-mediated PCR and 454 sequencing were all performed as described in Methods.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 29 October 2010; accepted 6 May 2011.

Published online 26 June 2011.

1. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nature Rev. Genet.* **11**, 636–646 (2010).

2. Urnov, F. D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646–651 (2005).
3. Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763 (2003).
4. Bibikova, M. *et al.* Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell. Biol.* **21**, 289–297 (2001).
5. Cartier, N. *et al.* Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science* **326**, 818–823 (2009).
6. Aiuti, A. *et al.* Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *N. Engl. J. Med.* **360**, 447–458 (2009).
7. Cideciyan, A. V. *et al.* Human gene therapy for RPE65 isomerase deficiency activates the retinoid cycle of vision but with slow rod kinetics. *Proc. Natl Acad. Sci. USA* **105**, 15112–15117 (2008).
8. Maguire, A. M. *et al.* Safety and efficacy of gene transfer for Leber's congenital amaurosis. *N. Engl. J. Med.* **358**, 2240–2248 (2008).
9. Bainbridge, J. W. *et al.* Effect of gene therapy on visual function in Leber's congenital amaurosis. *N. Engl. J. Med.* **358**, 2231–2239 (2008).
10. Cavazzana-Calvo, M. *et al.* Transfusion independence and *HMGA2* activation after gene therapy of human  $\beta$ -thalassaemia. *Nature* **467**, 318–322 (2010).
11. Howe, S. J. *et al.* Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J. Clin. Invest.* **118**, 3143–3150 (2008).
12. Hacein-Bey-Abina, S. *et al.* LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**, 415–419 (2003).
13. Hanna, J. *et al.* Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science* **318**, 1920–1923 (2007).
14. Pollak, E. S. & High, K. A. in *The Metabolic and Molecular Bases of Inherited Disease* (eds Scriver, C. R., Beaudet, A. L., Valle, D. & Sly, W. S.) 4393–4413 (McGraw-Hill, 2001).
15. Green, P. *Haemophilia B Mutation Database* (<http://www.kcl.ac.uk/ip/petergreen/haemBdatabase.html>) (2004).
16. Moehle, E. A. *et al.* Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc. Natl Acad. Sci. USA* **104**, 3055–3060 (2007).
17. Manno, C. S. *et al.* Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. *Nature Med.* **12**, 342–347 (2006).
18. Miao, C. H. *et al.* Inclusion of the hepatic locus control region, an intron, and untranslated region increases and stabilizes hepatic factor IX gene expression *in vivo* but not *in vitro*. *Mol. Ther.* **1**, 522–532 (2000).
19. Thompson, A. R. & Chen, S. H. Germ line origins of de novo mutations in hemophilia B families. *Hum. Genet.* **94**, 299–302 (1994).
20. Zambrowicz, B. P. *et al.* Disruption of overlapping transcripts in the ROSA  $\beta$ geo 26 gene trap strain leads to widespread expression of  $\beta$ -galactosidase in mouse embryos and hematopoietic cells. *Proc. Natl Acad. Sci. USA* **94**, 3789–3794 (1997).
21. Lin, H. F., Maeda, N., Smithies, O., Straight, D. L. & Stafford, D. W. A coagulation factor IX-deficient mouse model for human hemophilia B. *Blood* **90**, 3962–3966 (1997).
22. Perez, E. E. *et al.* Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nature Biotechnol.* **26**, 808–816 (2008).
23. Nakai, H. *et al.* Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction *in vivo*. *J. Virol.* **75**, 6969–6976 (2001).
24. Li, H. *et al.* Assessing the potential for AAV vector genotoxicity in a murine model. *Blood* **117**, 3311–3319 (2011).
25. Nakai, H. *et al.* AAV serotype 2 vectors preferentially integrate into active genes in mice. *Nature Genet.* **34**, 297–302 (2003).
26. Miller, D. G., Petek, L. M. & Russell, D. W. Adeno-associated virus vectors integrate at chromosome breakage sites. *Nature Genet.* **36**, 767–773 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was funded by the National Institutes of Health and the Howard Hughes Medical Institute.

**Author Contributions** H.L., V.H., Y.D., T.L., S.L.M., P.D.G., M.C.H. and K.A.H. designed the experiments. H.L., V.H., Y.D., T.L., S.Y.W., A.S.B., N.M., X.M.A., R.S., L.I., S.L.M., J.D.F., F.R.K., S.Z., D.E.P. and E.J.R. generated reagents and performed the experiments. H.L., Y.D., F.D.B., P.D.G., M.C.H. and K.A.H. wrote and edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to K.A.H. ([high@email.chop.edu](mailto:high@email.chop.edu)) or M.C.H. ([mholmes@sangamo.com](mailto:mholmes@sangamo.com); requests for zinc-finger nuclease reagents described in the paper).



## METHODS

**ZFN reagents.** ZFNs targeting the hF9 gene were designed by modular assembly using an archive of zinc finger proteins, as previously described<sup>3</sup>. The full amino acid sequences of the F9 ZFN pair are in Supplementary Fig. 1. The ZFN expression vector that was used *in vitro* was assembled as previously described<sup>27</sup>. The F9 ZFN AAV production plasmid was constructed by transferring the coding sequence into pRS115, a vector containing the AAV2 ITRs. ZFN expression was under the control of the ApoE enhancer and h $\alpha$ 1AT promoter from the previously described pAAV-hFIX16 plasmid<sup>17</sup>.

**Targeting vectors.** The NheI RFLP donor plasmid was constructed by amplifying 1-kb regions flanking the ZFN cleavage site from K-562-cell genomic DNA. A short sequence containing the NheI restriction site was subsequently introduced between the left and right arms of homology, as described in ref. 16. The NotI RFLP donor plasmid was constructed by amplifying the left (1 kb) and right (0.6 kb) arms of homology flanking the ZFN cleavage site from hF9mut mouse genomic DNA and cloning these into the production plasmid that contains the AAV2 ITRs, pRS165. A short sequence containing the NotI restriction site was subsequently introduced between the left and right arms of homology, as previously described<sup>16</sup>. The targeting vector used *in vivo* was built by cloning a cassette containing the splice acceptor, the coding sequence of exons 2–8 and the bovine growth hormone polyA signal from the pAAV-hFIX16 plasmid<sup>20</sup> into the NotI RFLP donor plasmid.

**Cell culture and transfection.** K-562 cells (ATCC) were maintained at 37 °C under 5% CO<sub>2</sub> in RPMI medium supplemented with 10% FBS, and were transfected using the 96-well Nucleofector kit SF (Lonza) as per the manufacturer's recommendations. Hep3B cells (ATCC) were maintained at 37 °C under 5% CO<sub>2</sub> in DMEM medium supplemented with 10% FBS, and were transfected using the 96-well Nucleofector kit SE (Lonza). HEK 293T cells (ATCC) were maintained at 37 °C under 5% CO<sub>2</sub> in DMEM medium supplemented with 10% FBS, and were transfected using the 96-well Nucleofector kit SF (Lonza). Lentiviral vector for stable transduction of the hF9mut mini-gene into HEK 293T cells was made using the ViraPower HiPerform lentiviral expression system (Invitrogen).

**Surveyor nuclease (Cel-I) assay and target-site sequencing.** Genomic DNA from K-562 and Hep3B cells was extracted using the QuickExtract DNA extraction solution (Epicentre Biotechnologies). ZFN target loci were amplified by PCR (30 cycles, 60 °C annealing and 30 s elongation at 68 °C) using the hF9cel1 forward primer (TCGGTGAGTGATTTGCTGAG) and hF9cel1 reverse primer (AACCTCTCACCTGGCCTCAT). Genomic DNA from mouse liver was isolated using the MasterPure complete DNA purification kit (Epicentre Biotechnologies). Primers for Cel-I of the hF9mut construct were hF9mut-cel1 forward (CTAGTAGCTGACAGTACC) and hF9mut-cel1 reverse (GAAGAACAGAAGCCTAATTA TG). The locus was amplified for 30 cycles (50 °C annealing and 30 s elongation at 68 °C). The assays were carried out as described previously<sup>22</sup>. For target-site sequencing, amplicons were cloned into the PCR-TOPO vector (Invitrogen) and sequenced using the primers M13forward (GTAAAACGACGGCCAGT) and M13reverse (GGAAACAGCTATGACCATG).

**RFLP knock-in and targeting assays.** Genomic DNA was extracted from K-562 and Hep3B cells using QuickExtract DNA extraction solution (Epicentre Biotechnologies). Genomic DNA from mouse liver was isolated using the MasterPure complete DNA purification kit (Epicentre Biotechnologies). The hF9 locus was amplified by 25 cycles of PCR (3 min extension at 68 °C and 30 s annealing at 55 °C) in the presence of radiolabelled dNTPs, using the hF9-TI forward (GGCCTTATTTACACAAAAGTCTG) and hF9-TI reverse (TTTGC TCTAACTCCTGTATCCATC) primers. The PCR products were then purified with G50 columns, digested with NheI, resolved by 5% PAGE and autoradiographed. RFLP assays in HEK 293T cells transduced with the hF9mut mini-gene were genotyped as described above, using the P1 (ACGGTATCGATAAGCTTGATATCGAATTCTAG) and P2 (CACTGATCTCCATCAACATACTGC) primers, and the PCR products (25 cycles, 63 °C annealing and 2 min extension at 65 °C) were digested with NotI. To quantify the targeting of the 'splice acceptor – exons 2–8 coding sequence – bovine growth hormone polyA signal' cassette, gDNA was amplified using the P1 and P3 (GAATAATTCTTTAGTTTGA GCAA) or the P1 and P2 primer pairs by 25 cycles of PCR (4 min extension time at 65 °C and 30 s annealing at 48 °C) in the presence of radiolabelled dNTPs. The PCR products were then purified with G50 columns, resolved by 5% PAGE and autoradiographed. All PCR reactions were performed using Accuprime Taq HiFi (Invitrogen). To capture the NHEJ-mediated insertion of the AAV vector at the hF9 ZFN cut-site, gDNA was amplified using P1 and P4 (AGGAACCCTAGTGATGGAG) primers by 25 cycles of PCR (80 s extension time at 65 °C) in the presence of radiolabelled dNTPs. The PCR reactions were performed using Phusion High-fidelity DNA polymerase (New England Biolabs) in conjunction with GC Buffer and 3% dimethylsulphoxide. The PCR products were then purified with G50 columns, resolved by 5% PAGE and autoradiographed.

**hF9mut mouse generation.** The hF9mut construct (sequence provided in Supplementary Fig. 14) was constructed by gene synthesis (Genscript) and ligated into the pUC57 plasmid. The hF9mut construct was then excised and ligated into a proprietary plasmid between FLP recombinase sites compatible for recombinase-mediated cassette exchange (RCME) (Taconic-Artemis), to create the hF9mut KI plasmid. The hF9mut KI plasmid and a FLP recombinase expression plasmid (Taconic-Artemis) were transfected into B6S6F1 embryonic stem (ES) cells (Taconic-Artemis) containing FLP recombinase sites compatible for RCME at the ROSA26 locus<sup>20</sup>. Correctly targeted B6S6F1-hF9mut ES cell clones were identified by Southern blot and injected into B6D2F1 blastocysts. Pure ES-cell-derived B6S6F1-hF9mut mice (G0) were delivered by natural birth and chimaeric pups were backcrossed with C57BL/6J mice (Jackson Laboratories) for 5 generations (for *in vivo* cleavage experiments) or 7–10 generations (for *in vivo* gene targeting experiments). hF9mut mice were genotyped using primers hF9mut Oligo 1 (ACTGTCTCTCATGCGTTGG), hF9mut Oligo 2 (GATGTTGGAGGTGGCA TGG), wtROSA Oligo 1 (CATGTCTTTAATCTACCTCGATGG), wtROSA Oligo 2 (CTCCCTCGTGATCTGCAACTCC), mFVIII Oligo1 (GAGCAAATTC CTGTACTGAC) and mFVIII Oligo 2 (TGCAAGGCCTGGGCTTATTT). HB mice have been backcrossed with C57BL/6J mice (Jackson Laboratories) for >10 generations and were genotyped using previously described primers<sup>20</sup>. C57BL/6J mice (Jackson Laboratories) were used for hF9mut-negative gene targeting experiments.

**AAV vector production.** AAV serotype 8 vectors were produced by triple transfection methods into HEK 293T cells, and subsequent CsCl density-gradient purification, as previously described<sup>28</sup>.

**Animal experiments.** AAV vector was diluted to 200  $\mu$ l with PBS before tail-vein injection. AAV vector was diluted to 20  $\mu$ l with PBS before neonatal intraperitoneal injection. Plasma for human factor IX ELISA was obtained by retro-orbital bleeding into heparinized capillary tubes. Plasma for aPTT was obtained by tail bleeding, 9:1 into 3.8% sodium citrate. Partial hepatectomies were performed as previously described<sup>29</sup>. Tissue for nucleic acid analysis was immediately frozen on dry ice after necropsy. All animal procedures were approved by the institutional animal care and use committee of the Children's Hospital of Philadelphia.

**SELEX.** *In silico* identification of potential off-target ZFN cleavage sites was performed by identifying homologous regions within the genome, as previously described<sup>22</sup>.

**LM-PCR and 454 sequencing.** AAV-donor integration junctions were cloned and sequenced as previously described<sup>24</sup>. In brief, genomic DNA from mouse liver was isolated using the MasterPure complete DNA purification kit (Epicentre Biotechnologies). 1  $\mu$ g of DNA was digested with MseI (New England Biolabs) and 1  $\mu$ g of DNA was digested with CviQ1 (New England Biolabs) for 16 h at 37 °C. These two enzymes were chosen for their ideal proximity to the target site. Digested DNA was purified using a PCR purification kit (Qiagen), then a previously described double-stranded linker<sup>24</sup> was ligated to digested DNA ends using T4 DNA ligase (New England Biolabs) for 16 h at 16 °C. Integration junctions were then PCR-amplified using an adaptor primer (GTAATACGACTCACTATAG GGC) and a stuffer primer (CTCCAATCTCTCAATCTCAGGTGATCTACCC). PCR products were diluted 1:200 in TE buffer and integration junctions were PCR-amplified again, using a second adaptor primer (CGTATCGCCTCCCTC GCGCCATCAGnnnnnnnnnnAGGGCTCCGCTTAAGGGAC, where nnnnnnnnnn is a sample-specific barcode) and a second stuffer primer (CTATGCGCCTTGCC AGCCGCTCAGnnnnnnnnnnACCTTGGCCTCCCAAATTGCTGGG, where nnnnnnnnnn is a sample-specific barcode). Amplified integration junctions were then sequenced using a Genome Sequencer FLX pyrosequencer (Roche/454).

**Integration-site analysis.** Pyrosequencing reads were first decoded using DNA barcodes, separating sequence reads by mouse. Reads were then aligned against the linker and stuffer primers using the Crossmatch program (-minmatch 8 -penalty -2 -minscore 6). Reads matching one or the other primer were then aligned using BLAT against three target sequences: the stuffer, the AAV-ITR and the hF9mut construct. BLAT parameters were optimized to find repetitive and/or short-sequence hits against each target sequence (-stepSize = 3, -tileSize = 8, -repMatch = 16384, -minScore = 5, -minIdentity = 50, -oneOff = 1). Additionally, BLAT fastMap option was included for alignment against the stuffer and the hF9mut construct. BLAT hits originating from the ITR were processed as previously described<sup>24</sup>. BLAT hits originating from the stuffer and the hF9mut construct were identified by requiring a unique high-scoring match requiring at least 90% sequence identity with a  $\leq 5$ -base-pair gap. All the BLAT hits from each of the three target sequences were consolidated and ordered by their location within each read. Reads that had stuffer and/or linker with ITR but no hF9mut construct were segregated and aligned using BLAT against the mouse genome. BLAT hits in the mouse genome were scored using the same criteria as described above and were required not to overlap with hits originating from stuffer, ITR and linker. A master table of all the reads and their respective target hits was constructed to manage the

alignment data and associated metadata. All the subsequent 454 analysis was carried out using this master table. Sequence analysis and control of mispriming was carried out separately for reads originating from each primer (stuffer or linker). To remove reads originating from mispriming at the stuffer primer, we required that each read involving the stuffer primer must extend through 30 base pairs (bp) of adjoining stuffer sequence and at least 13 bp of the flanking ITR. For reads originating on the linker side, we required that reads include at least 13 bp of ITR and at least 15 bases of the stuffer. Integration sites in the mouse genome were analysed as previously described<sup>30</sup>.

**Human factor IX quantification and functional analysis.** Quantification of human factor IX in plasma was performed using a human factor IX ELISA kit (Affinity Biologicals), with a standard curve from pooled normal human plasma (Trinity Biotech). All readings below the last value of the standard curve (15 ng ml<sup>-1</sup>) were arbitrarily given the value of 15 ng ml<sup>-1</sup>, the limit of detection. The assay of activated partial thromboplastin time (aPTT) was performed by mixing sample plasma 1:1:1 with pooled haemophilia B human plasma (George King

Biomedical, Inc.) and aPTT reagent (Trinity Biotech). Clot formation was initiated by addition of 25 mM calcium chloride.

**Liver function tests.** Quantification of plasma alanine aminotransferase (ALT) was performed using an ALT(SGPT) reagent set (Teco Diagnostics) colorimetric assay.

**Statistics.** Student's *t*-test was used as described. Linear regressions were performed using Prism (Graphpad). In all tests, differences were considered significant at  $P < 0.05$ .

27. Doyon, Y. *et al.* Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nature Biotechnol.* **26**, 702–708 (2008).
28. Ayuso, E. *et al.* High AAV vector purity results in serotype- and tissue-independent enhancement of transduction efficiency. *Gene Ther.* **17**, 503–510 (2010).
29. Mitchell, C. & Willenbring, H. A reproducible and well-tolerated method for 2/3 partial hepatectomy in mice. *Nature Protocols* **3**, 1167–1170 (2008).
30. Berry, C., Hannenhalli, S., Leipzig, J. & Bushman, F. D. Selection of target sites for mobile DNA integration in the human genome. *PLoS Comput. Biol.* **2**, e157 (2006).

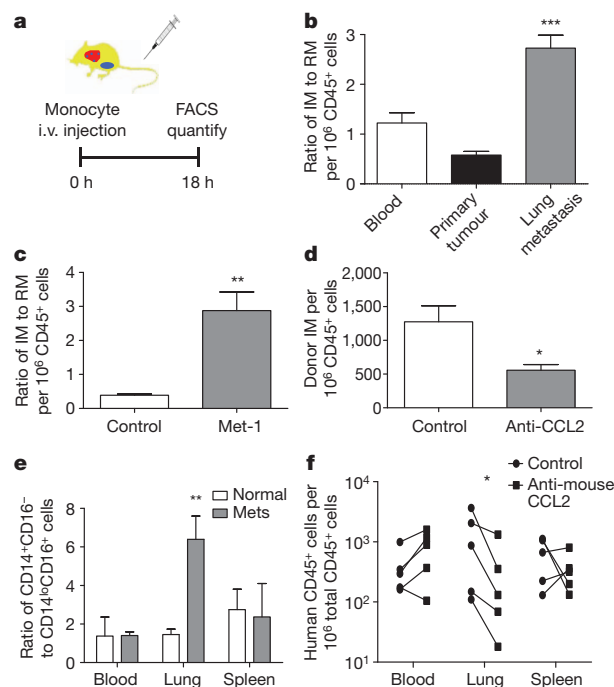
# CCL2 recruits inflammatory monocytes to facilitate breast–tumour metastasis

Bin-Zhi Qian<sup>1</sup>, Jiufeng Li<sup>1</sup>, Hui Zhang<sup>1</sup>, Takanori Kitamura<sup>1</sup>, Jinghang Zhang<sup>2</sup>, Liam R. Campion<sup>3</sup>, Elizabeth A. Kaiser<sup>3</sup>, Linda A. Snyder<sup>3</sup> & Jeffrey W. Pollard<sup>1</sup>

Macrophages, which are abundant in the tumour microenvironment, enhance malignancy<sup>1</sup>. At metastatic sites, a distinct population of metastasis-associated macrophages promotes the extravasation, seeding and persistent growth of tumour cells<sup>2</sup>. Here we define the origin of these macrophages by showing that Gr1-positive inflammatory monocytes are preferentially recruited to pulmonary metastases but not to primary mammary tumours in mice. This process also occurs for human inflammatory monocytes in pulmonary metastases of human breast cancer cells. The recruitment of these inflammatory monocytes, which express CCR2 (the receptor for chemokine CCL2), as well as the subsequent recruitment of metastasis-associated macrophages and their interaction with metastasizing tumour cells, is dependent on CCL2 synthesized by both the tumour and the stroma. Inhibition of CCL2–CCR2 signalling blocks the recruitment of inflammatory monocytes, inhibits metastasis *in vivo* and prolongs the survival of tumour-bearing mice. Depletion of tumour-cell-derived CCL2 also inhibits metastatic seeding. Inflammatory monocytes promote the extravasation of tumour cells in a process that requires monocyte-derived vascular endothelial growth factor. CCL2 expression and macrophage infiltration are correlated with poor prognosis and metastatic disease in human breast cancer<sup>3–6</sup>. Our data provide the mechanistic link between these two clinical associations and indicate new therapeutic targets for treating metastatic breast cancer.

To understand the origin of macrophages in primary tumours and their metastatic sites, we measured monocyte trafficking. Mouse monocytes were identified by their expression of CD11b and CD115 (Supplementary Fig. 3a) and were sorted by fluorescence-activated cell sorting (FACS) into sub-populations of inflammatory monocytes expressing Gr1 and Ly6c and resident monocytes lacking Gr1 and Ly6c (refs 7, 8) (Supplementary Fig. 3b–d). Both populations had similar expression of GFP in *Csflr-GFP* transgenic mice (Supplementary Fig. 3b). We adoptively transferred<sup>9–11</sup> 10<sup>5</sup> cells of each population into syngeneic FVB mice bearing autochthonous late-stage Polyoma Middle T (PyMT) mammary tumours with spontaneous pulmonary metastases (Fig. 1a). Eighteen hours after adoptive transfer, we determined the ratio of recovered GFP-positive inflammatory monocytes (Supplementary Fig. 3e) to resident monocytes from the same donor, to measure their relative recruitment. This indicated that there were similar numbers of donor cells in the blood (showing equivalent availability), but that in the primary tumour, resident monocytes were preferentially recruited, whereas in pulmonary metastases, inflammatory monocytes were preferentially recruited, with more than three-fold enrichment (Fig. 1b). Consistent with this, a notable population of endogenous inflammatory monocytes was identified in metastasis-bearing lungs but not in normal lungs (Supplementary Fig. 4a). This preferential recruitment of inflammatory monocytes in the lung was not observed in 7-week-old PyMT mice bearing pre-metastatic mammary tumours (Supplementary Fig. 4b). In experimentally induced pulmonary foci of intravenously injected Met-1 cells (a PyMT-induced mouse mammary tumour cell line)<sup>12</sup>, inflammatory monocytes were

also preferentially recruited (Supplementary Fig. 4c). GFP-labelled cells were readily detectable in pulmonary metastases at least 5 d after transfer (data not shown) and within 2 d, a significant portion of them had differentiated into F4/80<sup>+</sup> CD11b<sup>+</sup> Gr1<sup>–</sup> metastasis-associated macrophages (MAMs)<sup>2</sup> that are not seen in normal lungs (Supplementary Fig. 4d). To test whether inflammatory monocytes were recruited early in the metastasis process, we transferred monocyte populations



**Figure 1 | Pulmonary metastases preferentially recruit inflammatory monocytes through CCL2.** **a**, Schematic for the adoptive transfer of monocytes into PyMT-tumour-bearing mice with pulmonary metastases. i.v., intravenous. **b**, Ratios of inflammatory monocytes (IM) to resident monocytes (RM) in different tissues of recipient mice bearing PyMT tumours and metastases.  $n = 6$ ; \*\*\*,  $P < 0.0001$ . **c**, Ratios of inflammatory monocytes to resident monocytes in control lungs and in lungs with Met-1 cells intravenously injected 7 h before measurement.  $n = 4$ ; \*\*,  $P = 0.0039$ . **d**, Relative numbers of donor inflammatory monocytes recruited in lungs challenged with Met-1 cells for 7 h, with control or anti-mouse CCL2 antibody treatment.  $n = 3$ ; \*,  $P = 0.045$ . **e**, Ratios of adoptively transferred CD14<sup>+</sup>CD16<sup>–</sup> and CD14<sup>low</sup>CD16<sup>–</sup> human monocytes recruited into the lungs of normal mice (open bars) and of mice challenged with 4173 cells that contain metastases (Mets: solid bars) for 7 h.  $n = 5$ ; \*\*,  $P = 0.0163$ . All bars show mean  $\pm$  s.e.m. **f**, Numbers of adoptively transferred human CD14<sup>+</sup>CD16<sup>–</sup> monocytes that migrated into different tissues of mice challenged with 4173 cells via intravenous injection, with control or anti-mouse CCL2 antibody treatment. Each line connects data from the same donor.  $n = 5$ ; \*,  $P = 0.016$ .

<sup>1</sup>Department of Developmental and Molecular Biology, Center for the Study of Reproductive Biology and Women's Health, Albert Einstein College of Medicine, New York, New York 10461, USA. <sup>2</sup>Flow Cytometry Core Facility, Albert Einstein College of Medicine, New York, New York 10461, USA. <sup>3</sup>Ortho Biotech Oncology R&D, 145 King of Prussia Road, Radnor, Pennsylvania 19087, USA.



7 h after intravenous injection of Met-1 cells, a time point before significant interaction between the tumour and macrophages, and before extravasation of tumour cells<sup>2</sup>. Compared to control lungs, the recruitment of inflammatory monocytes to tumour-cell-challenged lungs increased markedly, with the ratio of inflammatory monocytes to resident monocytes increasing more than fivefold (Fig. 1c). However, this preferential recruitment of inflammatory monocytes was not observed after intravenous injection with PBS or latex beads, as controls for injection and particle lodgement, respectively (Supplementary Fig. 4e and data not shown). Consistent with this early recruitment of inflammatory monocytes, MAMs expressing high levels of CCR2 were preferentially recruited to lungs 36 h after tumour-cell inoculation<sup>2</sup>. However, B cells and T cells, including Foxp3<sup>+</sup> regulatory T (T<sub>reg</sub>) cells, were not differentially recruited at this time (Supplementary Fig. 5a–c and data not shown). These data indicate that MAMs are derived from inflammatory monocytes that are specifically recruited early in the process of pulmonary metastasis, before other immune cells.

Distinct chemokine signals recruit inflammatory and resident monocytes<sup>7</sup>, with inflammatory monocytes responding to CCL2 (refs 10, 11). Lung metastases of PyMT tumours express CCL2 homogeneously, in contrast to its heterogeneous expression in primary tumours (Supplementary Fig. 6a–d), and inflammatory monocytes have high levels of CCR2 expression, whereas resident monocytes do not (Supplementary Fig. 6e). The neutralization of CCL2 using a CCL2-specific antibody<sup>13</sup> markedly inhibited both the recruitment of inflammatory monocytes to lungs challenged with metastatic tumour cells (Fig. 1d) and the increase in the number of MAMs at the metastatic site (Supplementary Fig. 4f). Other CCR2-expressing leukocytes (a sub-population of T cells) and also T<sub>reg</sub> cells were unaffected by anti-CCL2 antibody treatment in this model (Supplementary Fig. 5d). Furthermore, the preferential recruitment of inflammatory monocytes to the tumour-cell-challenged lung was completely abrogated during adoptive transfer of monocytes sorted from *Ccr2*-null mice (Supplementary Fig. 6f).

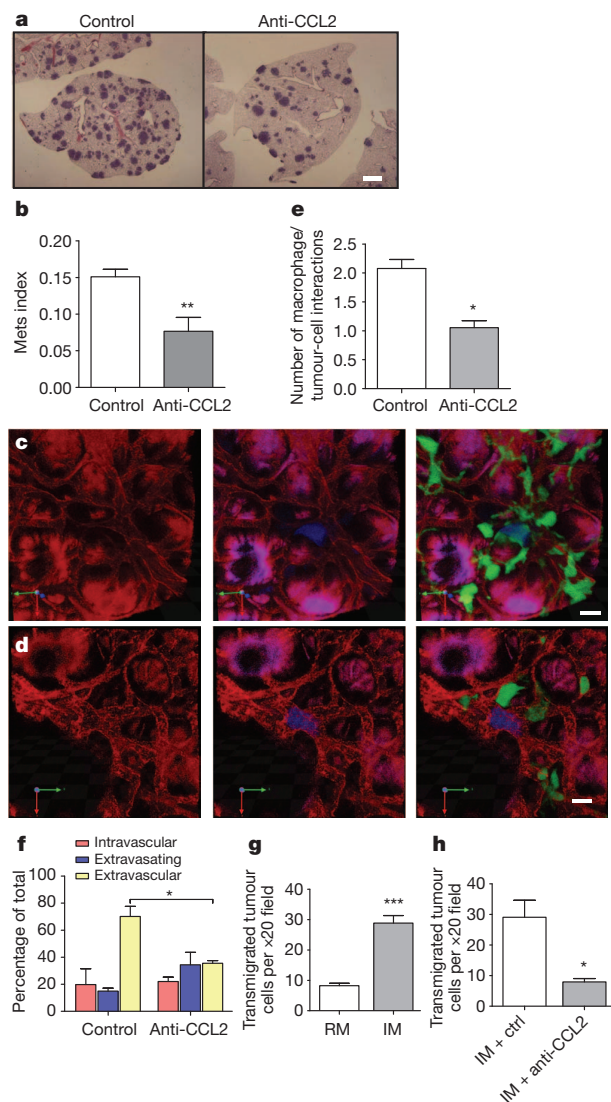
The pattern of human monocyte recruitment to tumours *in vivo* is unknown. To investigate this, human CD14<sup>+</sup>CD16<sup>−</sup> inflammatory monocytes and CD14<sup>low</sup>CD16<sup>+</sup> resident monocytes<sup>9</sup> were sorted from enriched CD14<sup>+</sup> cells from the peripheral blood of healthy donors (Supplementary Fig. 7a). 10<sup>5</sup> cells of each population were adoptively transferred into pairs of nude mice supplemented with recombinant human colony-stimulating factor 1 (CSF1), which is essential for the survival of monocytes and macrophages (Supplementary Fig. 7e). Human monocytes were quantified 18 h after adoptive transfer, using FACS analysis with an antibody against human CD45 (Supplementary Fig. 7b). In normal mice, after adoptive transfer of monocytes from the same donor, there were comparable numbers of human inflammatory monocytes and resident monocytes in the circulation and also recruited to the lung, but about twice the numbers of inflammatory monocytes compared to resident monocytes in the spleen (Fig. 1e, open bars). In mice given an intravenous injection of human MDA-MB-231-derived metastatic 4173 breast cancer cells<sup>14</sup> 7 h before monocyte transfer, the ratio of the two monocyte populations in blood and spleen was similar to that in normal mice, but the ratio of inflammatory monocytes to resident monocytes in the lungs increased more than sixfold (Fig. 1e). In established pulmonary metastases derived from orthotopically injected 4173 cells, inflammatory monocytes were also preferentially recruited, with a ratio fivefold higher than that in normal lungs (Supplementary Fig. 7d). Mouse inflammatory monocytes were also preferentially recruited to lungs challenged with 4173 cells (data not shown). Human inflammatory monocytes express CCR2, whereas resident monocytes express minimal levels of this receptor (Supplementary Fig. 7c). The neutralization of host CCL2 with an antibody against mouse CCL2 markedly reduced the recruitment of human inflammatory monocytes into lungs challenged with 4173 cells, without any change in the circulation or spleen (Fig. 1f). Treatment with an antibody specific to human CCL2 (ref. 15) also inhibited inflammatory monocyte recruitment (Supplementary Fig. 7f), indicating the importance of CCL2 from both the tumour and the

target organ. This shows that human inflammatory monocytes respond to the same CCL2–CCR2 signalling as mouse cells for their specific recruitment during pulmonary metastasis.

To test the effect on metastatic potential of blocking the recruitment of inflammatory monocytes, we performed experimental metastasis assays with Met-1 cells in mice treated with anti-mouse CCL2 or with a control antibody shortly before the tumour-cell injection. Anti-CCL2 treatment reduced the total metastasis burden, owing to a markedly reduced number of metastasis nodules (Fig. 2a, b). An antibody specific to mouse CCL12, another ligand of mouse CCR2, had no effect on the metastasis of Met-1 cells (Supplementary Fig. 8). This indicates that the specific CCL2-mediated recruitment of inflammatory monocytes is critical for the pulmonary seeding of tumour cells.

Extravasation is a critical step for the metastatic seeding of tumour cells in the lung<sup>2</sup>. We used an intact-lung imaging system<sup>16</sup> to test the role of CCL2-recruited inflammatory monocytes in tumour-cell extravasation. *Csf1r*-GFP transgenic mice were injected intravenously with cyan fluorescent protein (CFP)-expressing Met-1 cells and analysed after 24 h. Quantification of three-dimensional reconstructed confocal images (Fig. 2c, d and Supplementary Movies 1 and 2) showed that the number of macrophages interacting directly with tumour cells was significantly reduced by anti-mouse CCL2 neutralizing antibody, compared with control antibody (Fig. 2e). Notably, tumour-cell extravasation was delayed and less efficient after the blocking of inflammatory monocytes (Fig. 2f). Tumour-cell extravasation involves crosstalk between tumour cells, endothelial cells, basement membrane and macrophages. In an *in vitro* trans-endothelial migration assay (Supplementary Fig. 9a)<sup>17</sup>, the trans-endothelial migration of tumour cells was enhanced about fivefold by mouse bone-marrow-derived macrophages (BMDMs) located on the basolateral side of the endothelial monolayer. This effect was blocked by anti-mouse-CCL2 neutralizing antibody, but not by control antibody (Supplementary Fig. 9b). Tumour cells, BMDMs and endothelial cells all express CCL2, whereas only the macrophages express CCR2 (Supplementary Fig. 9c), indicating that only macrophages respond to the CCL2 chemokine signalling. In confirmation of this, macrophages from *Ccr2*-null mice were not capable of promoting trans-endothelial migration of tumour cells (Supplementary Fig. 9d). Notably, FACS-sorted inflammatory monocytes, but not resident monocytes, markedly promoted tumour-cell trans-endothelial migration and this was also inhibited by anti-mouse-CCL2 neutralizing antibody (Fig. 2g, h).

Total blockade of CCL2 (both mouse and human) inhibited spontaneous lung metastasis of orthotopically injected MDA-MB-231 cells (Fig. 3a). Ligands secreted by both the tumour cells and the host contributed to metastatic efficiency, because both anti-human and anti-mouse antibodies markedly inhibited the experimental metastasis of 4173 cells (Fig. 3b) without affecting tumour-cell proliferation *in vitro* (data not shown). This conclusion was also confirmed by knocking down CCL2 using small interfering RNAs in 4173 cells: this markedly reduced lung colonization in experimental metastasis assays (Supplementary Fig. 10e, f). Consistent with this, a similar *Ccl2*-knockdown in Met-1 cells did not affect tumour-cell proliferation *in vitro*, but markedly inhibited the metastatic efficiency of the cells (Supplementary Fig. 10a–c). Trans-endothelial migration of 4173 cells *in vitro* was also promoted by human inflammatory monocytes and inhibited by neutralizing either human or mouse CCL2 with specific antibodies (Fig. 3c–e). These data indicate that CCL2 secreted by both the tumour cell and the target organ promotes tumour-cell extravasation and metastatic seeding via the recruitment of inflammatory monocytes. Consistent with the role of CCL2 synthesized by the microenvironment in the lung, bone metastases of MDA-MB-231 cells also recruit inflammatory monocytes and the inhibition of CCL2 inhibits metastatic progression. In contrast, liver metastases of Met-1 cells did not recruit inflammatory monocytes and CCL2 inhibition did not reduce metastasis (data not shown). Furthermore, CCL2 blockade 2 d after intravenous injection of MDA-MB-231 cells reduced the tumour burden in the lung and prolonged the

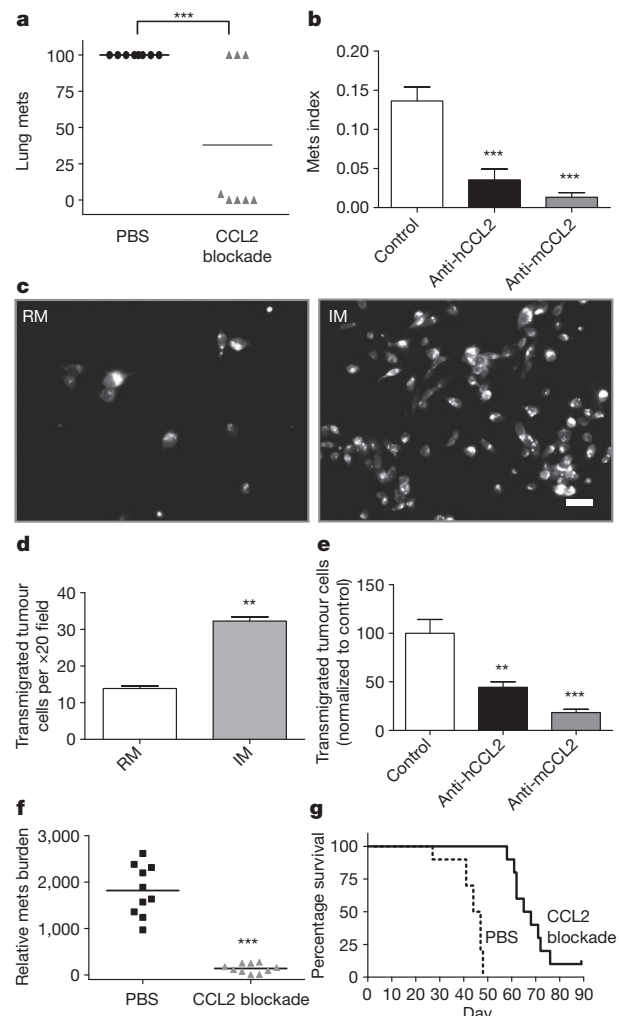


**Figure 2 | CCL2-recruited monocytes promote metastatic seeding.**

**a**, Representative haematoxylin-&-eosin-stained sections showing Met-1 metastasis with control or anti-CCL2 antibody treatment. Scale bar, 1 mm. **b**, Met-1 metastasis (Mets) burden with or without antibody treatment.  $n = 6$ ; \*\*,  $P = 0.006$ . **c, d**, Representative snapshots of three-dimensional reconstructed confocal images of tumour cells (blue) and macrophages (green) in lung vasculature (red) 24 h after tail-vein injection of tumour cells into mice treated with control (**c**) or anti-mouse CCL2 (**d**) antibodies. Scale bar, 20  $\mu$ m. Arrows define the dimensions of the figure. **e, f**, Numbers of interactions between macrophages and tumour cells (**e**) and tumour-cell extravasation (**f**) in mice with control or anti-mouse CCL2 antibody treatment. (**e**,  $P = 0.0066$ , and **f**,  $P = 0.00163$ , are based upon three-dimensional images of 15–20 tumour clusters per mouse,  $n = 3$  mice per group.) **g**, Numbers of transigrated Met-1 cells in the presence of resident monocytes or inflammatory monocytes.  $n = 5$ ; \*\*\*,  $P < 0.0001$ . **h**, Numbers of transigrated Met-1 cells in the presence of inflammatory monocytes, with antibody treatments.  $n = 3$ ; \*,  $P = 0.0204$ . All bars show mean + s.e.m.

survival of mice, indicating the importance of continuous recruitment of inflammatory monocytes and their differentiation into MAMs for persistent metastatic growth (Fig. 3f, g).

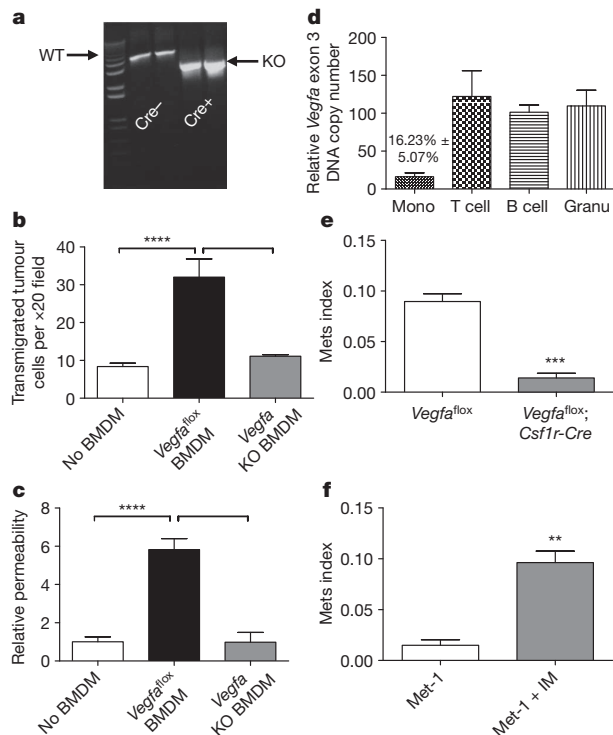
To determine a mechanism for the effects of inflammatory monocytes on tumour-cell extravasation, we analysed the transcriptomes of resident and inflammatory monocytes<sup>18</sup>. Among the differentially regulated genes, vascular endothelial growth factor A (*Vegfa*) was highly expressed by inflammatory monocytes, a fact that we verified experimentally (Supplementary Fig. 11a). To ablate *Vegfa* conditionally in myeloid cells to test its role in the metastatic process, we generated a transgenic mouse expressing a tamoxifen-inducible Mer-iCre fusion protein driven by the



**Figure 3 | CCL2 from both the tumour cell and the host promotes metastatic seeding.**

**a**, Numbers of spontaneous pulmonary metastases from orthotopic MDA-MB-231 tumours with total CCL2 blockade or control treatment. Bar shows the mean;  $n = 8$  per group; \*\*\*,  $P < 0.001$ . **b**, Metastasis burden of intravenously injected 4173 cells with different antibody treatments. Bars show mean + s.e.m.;  $n = 6$ ; \*\*\*,  $P = 2.14 \times 10^{-5}$ . **c**, Representative fluorescent micrographs of transigrated human 4173 cells pre-stained with cell-tracker dye in the presence of inflammatory or resident monocytes. Scale bar, 20  $\mu$ m. **d**, Numbers of transigrated 4173 cells in the presence of inflammatory monocytes or resident monocytes. Bars show the mean + s.e.m. of three experiments with duplicates; \*\*,  $P = 0.0051$ . **e**, Relative number of transigrated 4173 cells in the presence of inflammatory monocytes with control, anti-human CCL2 or anti-mouse CCL2 antibodies, normalized to the average number with control antibody treatment, which is set to 100. Bars represent the mean + s.e.m. of five experiments with duplicates. One-way analysis of variance with Bonferroni's multiple comparison test; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ . **f, g**, CCL2 blockade starting from 2 d after intravenous injection of MDA-MB-231 cells significantly reduces the metastasis burden, as measured by real-time PCR of human *Alu* repeats, normalized to mouse  $\beta$ -actin, on day 22 (**f**,  $n = 10$ ; \*\*\*,  $P < 0.001$ ). CCL2 blockade also prolongs survival (**g**,  $n = 10$ ,  $P < 0.001$ ) compared to control treatment with PBS.

*Csf1r* promoter, crossed with *Vegfa*<sup>flox/flox</sup> mice<sup>19</sup>. Inducible ablation of *Vegfa* was achieved in cultured BMDMs treated with 4-hydroxytamoxifen (Fig. 4a) and these *Vegfa*-null BMDMs were unable to promote the trans-endothelial migration of tumour cells and did not enhance permeability of the endothelial monolayer, a process important for metastasis<sup>20</sup>, when compared to control BMDMs (Fig. 4b, c). *In vivo* injection of tamoxifen specifically ablated *Vegfa* in monocytes, without ablation in other circulating immune cells (Fig. 4d). This monocyte-specific depletion of VEGFA markedly inhibited the potential for experimental



**Figure 4 | Monocyte-specific ablation of *Vegfa* blocks pulmonary seeding.** **a**, PCR of *Vegfa* exon 3 in BMDMs from *Vegfa*<sup>flox/flox</sup> mice, with or without the *Csf1r-Mer-iCre-Mer* transgene, treated with 4-hydroxytamoxifen. Wild-type (WT) and knockout (KO) bands are indicated. **b**, **c**, Numbers of trans-endothelial migrated Met-1 cells (**b**) and permeability of the endothelial monolayer to albumin (**c**), with no BMDMs, *Vegfa*<sup>flox</sup> BMDMs or *Vegfa*-knockout BMDMs. *n* = 3 with duplicates; \*\*, *P* < 0.01 with analysis of variance. **d**, Relative copy number of *Vegfa* exon 3 in leukocytes from the peripheral blood of tamoxifen-treated *Vegfa*<sup>flox/flox</sup> *Csf1r-Mer-iCre-Mer* mice compared with *Vegfa*<sup>flox/flox</sup> mice. Mono, monocyte; granu, granulocyte. **e**, Met-1 Mets burden in *Vegfa*<sup>flox/flox</sup> mice with or without *Cre*, with the same tamoxifen treatment. *n* = 6; \*\*\*, *P* = 0.0004. **f**, Met-1 Mets burden in *Vegfa*<sup>flox/flox</sup> *Csf1r-Mer-iCre-Mer* mice with tamoxifen treatment, with or without co-injection of inflammatory monocytes. *n* = 6; \*\*, *P* < 0.0001. All data are mean + s.e.m.

metastasis of Met-1 cells and reduced their seeding efficiency (Fig. 4e and Supplementary Fig. 11b). Adoptive transfer experiments indicated that *Vegfa*-null inflammatory monocytes infiltrate Met-1 lung metastases at a comparable level to *Vegfa*<sup>flox</sup> inflammatory monocytes, showing that VEGFA is not required for the recruitment of these cells (Supplementary Fig. 11c). Notably, co-injection of Met-1 cells and wild-type inflammatory monocytes into inducible macrophage-*Vegfa*-knockout mice restored the metastatic potential of tumour cells (Fig. 4f).

These experiments indicate that CCL2 synthesized by metastatic tumour cells and by the target-site tissue stroma is critical for the recruitment of a sub-population of CCR2-expressing monocytes that enhance the subsequent extravasation of the tumour cells. Mechanistically, this occurs at least in part through targeted delivery of molecules such as VEGFA that promote extravasation. Inflammatory monocytes are continually recruited by a CCL2-dependent mechanism and differentiate into macrophages that promote the subsequent growth of metastatic cells (Supplementary Fig. 1). These data, together with the clinical association of CCL2 overexpression in human cancers with poor prognosis (Supplementary Fig. 2), strongly argue for therapeutic approaches targeted against monocyte recruitment and function.

## METHODS SUMMARY

The trafficking of monocytes into primary tumours and their metastases was studied by adoptive transfer of mouse (Ly6c/Gr1<sup>+</sup> or Ly6c/Gr1<sup>-</sup>) monocytes or human (CD14<sup>+</sup>CD16<sup>+</sup> and CD16<sup>-</sup>) monocytes, using MMTV-PyMT autochthonous,

human and mouse experimental metastasis models and human orthotopic tumour models. Monocytes and macrophages were recovered by enzymatic disaggregation of the tumours, followed by FACS analysis. To investigate mechanisms for monocyte recruitment and the effect of inhibition of this recruitment on metastasis, anti-mouse-CCL2 or anti-human-CCL2 antibodies or *Ccr2*-null mutant mice were used. To ablate *Vegfa* expression in monocytes, a myeloid-specific (*Csf1r* promoter), tamoxifen-inducible *Cre*-expressing strain was crossed with *Vegfa*<sup>flox/flox</sup> mice and gene ablation was induced by tamoxifen. The effect of monocyte depletion on tumour-cell extravasation using Met-1, an FVB PyMT-tumour-derived metastatic cell line, was determined using an *ex vivo* intact-lung imaging system and an *in vitro* extravasation assay.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 24 May 2010; accepted 19 April 2011.

Published online 8 June 2011.

- Qian, B. Z. & Pollard, J. W. Macrophage diversity enhances tumor progression and metastasis. *Cell* **141**, 39–51 (2010).
- Qian, B. *et al.* A distinct macrophage population mediates metastatic breast cancer cell extravasation, establishment and growth. *PLoS ONE* **4**, e6562 (2009).
- Ueno, T. *et al.* Significance of macrophage chemoattractant protein-1 in macrophage recruitment, angiogenesis, and survival in human breast cancer. *Clin. Cancer Res.* **6**, 3282–3289 (2000).
- Valkovic, T., Lucin, K., Krstulja, M., Dobi-Babic, R. & Jonjic, N. Expression of monocyte chemotactic protein-1 in human invasive ductal breast cancer. *Pathol. Res. Pract.* **194**, 335–340 (1998).
- Saji, H. *et al.* Significant correlation of monocyte chemoattractant protein-1 expression with neovascularization and progression of breast carcinoma. *Cancer* **92**, 1085–1091 (2001).
- Rhodes, D. R. *et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1–6 (2004).
- Geissmann, F. *et al.* Blood monocytes: distinct subsets, how they relate to dendritic cells, and their possible roles in the regulation of T-cell responses. *Immunol. Cell Biol.* **86**, 398–408 (2008).
- Geissmann, F. *et al.* Development of monocytes, macrophages, and dendritic cells. *Science* **327**, 656–661 (2010).
- Geissmann, F., Jung, S. & Littman, D. R. Blood monocytes consist of two principal subsets with distinct migratory properties. *Immunity* **19**, 71–82 (2003).
- Getts, D. R. *et al.* Ly6c<sup>+</sup> “inflammatory monocytes” are microglial precursors recruited in a pathogenic manner in West Nile virus encephalitis. *J. Exp. Med.* **205**, 2319–2337 (2008).
- Palframan, R. T. *et al.* Inflammatory chemokine transport and presentation in HEV: a remote control mechanism for monocyte recruitment to lymph nodes in inflamed tissues. *J. Exp. Med.* **194**, 1361–1374 (2001).
- Borowsky, A. D. *et al.* Syngeneic mouse mammary carcinoma cell lines: two closely related cell lines with divergent metastatic behavior. *Clin. Exp. Metastasis* **22**, 47–59 (2005).
- Tsui, P. *et al.* Generation, characterization and biological activity of CCL2 (MCP-1/JE) and CCL12 (MCP-5) specific antibodies. *Hum. Antibodies* **16**, 117–125 (2007).
- Minn, A. J. *et al.* Genes that mediate breast cancer metastasis to lung. *Nature* **436**, 518–524 (2005).
- Carton, J. M. *et al.* Codon engineering for improved antibody expression in mammalian cells. *Protein Expr. Purif.* **55**, 279–286 (2007).
- Al-Mehdi, A. B. *et al.* Intravascular origin of metastasis from the proliferation of endothelium-attached tumor cells: a new model for metastasis. *Nature Med.* **6**, 100–102 (2000).
- Ma, C. & Wang, X.-F. *In vitro* assays for the extracellular matrix protein-regulated extravasation process. *CSH Protoc.* doi:10.1101/pdb.prot5034 (2008).
- Swirski, F. K. *et al.* Identification of splenic reservoir monocytes and their deployment to inflammatory sites. *Science* **325**, 612–616 (2009).
- Gerber, H. P. *et al.* VEGF is required for growth and survival in neonatal mice. *Development* **126**, 1149–1159 (1999).
- Huang, Y. *et al.* Pulmonary vascular destabilization in the premetastatic phase facilitates lung metastasis. *Cancer Res.* **69**, 7529–7537 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by grants from the NIH to J.W.P. (NIH P01 CA100324 and R01 CA131270) and to the Albert Einstein Cancer Center Core (P30 CA 13330). We thank J. Massague for 4173 cells and N. Ferrara for the *Vegfa*<sup>flox/flox</sup> mice. We also thank P. Marsters for statistical analyses and M. Thompson, F. Shi, C. Ferrante, F. McCabe, H. Millar-Quinn and D. Wiley for discussions and technical assistance.

**Author Contributions** B.-Z.Q., L.A.S. and J.W.P. conceived the ideas and designed the experiments. B.-Z.Q., J.L., H.Z., T.K., J.Z., L.R.C. and E.A.K. performed the experiments. B.-Z.Q., J.L., L.A.S. and J.W.P. analysed the data. B.-Z.Q., L.A.S. and J.W.P. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.W.P. ([pollard@aeom.yu.edu](mailto:pollard@aeom.yu.edu)).



## METHODS

**Animals.** All procedures involving mice were conducted in accordance with National Institutes of Health regulations concerning the care and use of experimental animals. The study of mice was approved by the Albert Einstein College of Medicine and Ortho Biotech R&D Institute animal care and use committees. Transgenic mice expressing the Polyoma Middle T (PyMT) oncogene under the control of the mouse mammary tumour virus long terminal repeat (MMTV LTR) promoter were provided by W. J. Muller and were bred in-house. FVB (*Tg(Csf1r-EGFP)1Jwp*) mice have been previously reported to have the whole mononuclear phagocyte system labelled<sup>2</sup>. BL6 *Ccr2*<sup>tm1lfc/J</sup> mice were purchased from The Jackson Laboratory. The FVB macrophage-specific (*Csf1r* promoter), tamoxifen-inducible Cre-expressing *Tg(Csf1r-Mer-iCre-Mer)1Jwp* transgenic mouse strain was generated and crossed with *Vegfa*<sup>fllox/fllox</sup> mice (gift from N. Ferrara). Knockout of *Vegfa* in myeloid cells was induced by daily subcutaneous injection of 3 µg tamoxifen per mouse for 2 d, before sorting for blood leukocytes or tumour-cell injection.

**Metastasis assay.** Eight-week-old FVB females and six-week-old female nude mice were used for lung experimental-metastasis assays with intravenous injection of  $5 \times 10^5$  Met-1 cells or  $10^6$  MDA-MB-231-derived LM2 human breast cancer cells, 4173 (ref. 14), respectively. If not otherwise specified, all animals were killed 2 weeks after intravenous injection of Met-1 cells or 4 weeks after injection of human tumour cells, for optimal metastatic burden. In experimental metastasis assays, antibodies were given at  $10 \text{ mg kg}^{-1}$  body weight via intraperitoneal injection 3 h before tumour-cell injection, for single treatments, or twice a week thereafter for prolonged treatments, if not otherwise specified. For paraffin sections, lungs were injected with 1.2 ml of 10% neutral buffered formalin by tracheal cannulation to fix the inner airspaces and inflate the lung lobes. Lungs were excised and fixed in formalin overnight. A precise stereological method<sup>21</sup> with modification was used for quantification of lung metastases. Briefly, paraffin-embedded lungs were systematically sectioned through the entire lung with one 5 µm section taken in every 0.5 mm of lung thickness. All sections were stained with haematoxylin and eosin and images were taken using a Zeiss SV11 microscope with a Retiga 1300 digital camera and analysed using ImageJ<sup>22</sup>. The Mets index is the total volume of metastases normalized to total lung volume and Mets number is the number of metastasis nodules per  $\text{mm}^2$  of lung area. Real-time PCR quantification of the burden of human tumour cells was performed as reported previously, using human-specific primers<sup>23</sup>. For spontaneous lung metastasis,  $2.5 \times 10^6$  parental MDA-MB-231 cells or  $10^6$  derived LM2 4173 tumour cells were orthotopically injected into the inguinal mammary gland of SCID beige or nude mice, respectively. Anti-mouse CCL2 and CCL12 (ref. 23) and anti-human CCL2 (ref. 15) antibodies neutralize only their respective target molecules and were provided by Ortho Biotech Oncology together with the control antibody. In spontaneous metastasis assays, antibody treatment began on day 3 after tumour-cell intra-mammary-gland injection and continued twice a week thereafter, with each antibody used at  $20 \text{ mg kg}^{-1}$  body weight. When each group reached a mean primary-tumour volume of  $\sim 1,000 \text{ mm}^3$ , the mice were killed. Lungs were perfused with India ink and placed in Fekete's solution. Lung metastases were counted in a blinded fashion. All *in vivo* experiments were at least two independent experiments with 3–10 mice for each group.

**Adoptive transfer.** CD115<sup>+</sup> F4/80<sup>+</sup> CD11b<sup>+</sup> Ly6c1/Gr1<sup>+</sup> and Ly6c1/Gr1<sup>−</sup> bone-marrow monocytes were sorted from FVB *Csf1r*-EGFP mice and adoptively transferred into FVB mice.  $10^5$  of either cell type were transferred into mice bearing mammary tumours and/or pulmonary metastases. Monocytes were sorted from *Ccr2*-null mutant mice using the same protocol and were labelled with CellTracker (Invitrogen) following the manufacturer's instructions, before adoptive transfer into nude mice. Fresh human CD14<sup>+</sup> peripheral monocytes were purchased from All Cells LLC.  $10^5$  CD14<sup>+</sup>CD16<sup>−</sup> and CD14<sup>+</sup>CD16<sup>+</sup> cells were FACS-sorted and intravenously transferred into nude mice supplemented with  $2 \times 10^6$  units of recombinant human CSF1 via subcutaneous injection. In the indicated experiments, specified antibodies were given at  $10 \text{ mg kg}^{-1}$  body weight 3 h before adoptive transfer of monocytes.

**FACS analysis and antibodies.** For FACS analysis, lungs or whole mice were perfused thoroughly with cold PBS before cell collection, then lungs were minced on ice and digested with an enzyme mix of Liberase and Dispase (Invitrogen). Blood was drawn by cardiac puncture. Red blood cells were removed using RBC lysis buffer (eBioscience). Cells were blocked using anti-mouse CD16/CD32 antibody (eBioscience) for mouse cells, or 10% goat serum for human cells, before antibody staining. Antibodies against mouse antigens were: CD45 (30-F11), CD11b (M1/70), Gr1 (RB6-8C5), CD115 (AFS98) and Foxp3 (FJK-16 s; all from eBioscience); CD3 (145-2C11) and Ly6c1 (HK1.4; both from Biolegend); CD25 (PC61), CD62L (MEL-14), IL4Ra (mIL4R-M1), CD4 (GK1.5), CD8a (53-6.7) and Ly6G (1A8; all from BD Pharmingen) and F4/80 (Cl:A3-1; AbD Serotec). Antibodies against human antigens were: CD14 (Tük4) and CD16 (3G8; both from Invitrogen), CD45 (HI30; BioLegend) and CCR2 (48607; R&D Systems). FACS analysis was performed on a LSRII cytometer (BD Biosciences) and data

were analysed using Flowjo software (TreeStar). Gating of single cells using FSC/W and SSC/W and exclusion of dead cells with DAPI staining were performed routinely during analysis. Mouse CCL2 was stained using the specific antibody R-17 (Santa Cruz) after a standard immunohistochemistry protocol.

**Cell culture and *in vitro* extravasation assay.** All cells were cultured in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal bovine serum (FBS). The extravasation assay was performed as previously described<sup>17,24</sup> with modifications. Briefly,  $2 \times 10^4$  endothelial cells (3B-11, ATCC) were plated into the upper chamber of a GFR matrigel invasion chamber (BD Biosciences) in DMEM with 10% (v/v) FBS. A monolayer was formed in 2 d and was verified by microscopy.  $10^4$  BMDMs or FACS-sorted monocytes were loaded to the basolateral side of the insert and put into a plate-well with DMEM, 10% FBS and  $10^4$  units  $\text{ml}^{-1}$  CSF1 to allow attachment. *Vegfa*-knockout BMDMs derived from *Csf1r-Mer-iCre-Mer-Vegfa*<sup>fllox/fllox</sup> mice were induced by treating the cells with  $1 \mu\text{M}$  4-hydroxyltamoxifen for 7 d after isolation of bone marrow.  $2 \times 10^4$  Met-1 cells stained with CellTracker CMRA (Invitrogen) were loaded into the insert with DMEM in 0.5% (v/v) FBS and  $10^4$  units  $\text{ml}^{-1}$  CSF1. CCL2-neutralizing antibody and control antibody were used at  $5 \mu\text{g ml}^{-1}$ , applied to both sides of the insert. Plates were incubated under normal tissue-culture conditions for 36–48 h before being fixed with 1% (w/v) paraformaldehyde. Tumour-cell trans-endothelial migration was quantified by counting the number of cells that migrated through the insert under a fluorescent microscope (6–10 randomly-selected fields in each insert) and was expressed as cell number per  $\times 20$  field, if not otherwise specified. The permeability assay was performed by loading 4% (w/v) bovine serum albumin labelled with Evan's blue into the upper chamber with a pre-formed endothelial monolayer of 3B-11 cells and measuring the absorption of the phenol-red-free medium in the lower chamber at 650 nm after a 30-min incubation in normal culturing conditions. All *in vitro* experiments were at least three independent experiments with duplicate or triplicate measures.

**Molecular biology.** To knockdown *Ccl2* in Met-1 cells, a 97-mer oligo containing a small hairpin RNA (shRNA) that targets the *Ccl2* mRNA sequence from nucleotide 166 was cloned into the miR30 context in the retroviral vector P2GM<sup>25</sup>. To knockdown *CCL2* in 4173 cells, a 97-mer oligo containing a shRNA targeting the human *CCL2* mRNA sequence from nucleotide 255 was cloned into the miR30 context in the same vector. For real-time PCR of mouse *Ccl2* expression, primers CCCAATGAGTAGGCTGGAGA and AAAATGGATC CACACCTTGC were used, and for *Ccr2*, primers CCTGCAAGACCAAGAAG AGG and GTGAGCAGGAAGAGCAGGTC. All real-time PCR was performed on an MJ Research DNA Engine 2 Opticon real-time PCR machine using SYBR master mix (Invitrogen). Primers used were: mouse *Ccl2* primers GTTGGC TCAGCCAGATGCA and AGCCTACTCATTGGGATCATCTTG; mouse *Ccr2* TTTGTTTTTGCAGATGATTCAA and TGCCATCATAAAGGAGCCAT; mouse *Plau* ACAGATAAGCGGTCTCCAG and GCCCACTACTATGGCTC TG; mouse *Vegfa* AATGCTTTCTCCGCTCTGAA and GCTTCCTACAGCACA GCAGA; mouse *Vegfa* exon 3 ACATCTTCAAGCCGCTCTGT and TGCAT GGTGATGTTGCTCT; human *CCL2* AGGTGACTGGGGCATTGAT and GCCTCCAGCATGAAAGTCTC. To verify mouse *Vegfa* exon 3 knockout, primers that flank this exon, GCTGCACCCACGACAGAAGG and TGAGGTT TGATCCGCATGAT, were used.

**Ex vivo whole-lung imaging.** A well-established intact-lung microscopy technique<sup>16,26</sup> was applied to observe tumour cells, macrophages and blood vessels in mouse lungs. CFP-expressing Met-1 cells, prepared by retrovirus infection of a CMV-promoter CFP vector, were injected intravenously into the tail vein of each mouse. At the times indicated, mice were anaesthetized and injected with  $10 \mu\text{g}$  AlexaFluor-647-conjugated anti-mouse CD31 antibody (BioLegend). Five minutes later, the mouse was put under artificial ventilation through tracheal cannulation. The lung was cleared of blood by gravity perfusion through the pulmonary artery with artificial medium (Kreb-Ringer bicarbonate buffer with 5% dextran and  $10 \text{ mmol l}^{-1}$  glucose (pH 7.4)). The heart-lung preparation was dissected en bloc and placed in a specially designed plexiglass chamber with a port to the artificial cannula. The lung rested on a plexiglass window at the bottom of the chamber with the posterior surface of the lung touching the plexiglass. The lung was ventilated throughout the experiment with 5%  $\text{CO}_2$  in medical air and perfused by gravity perfusion except during imaging. Three to five animals were imaged for each time point and 10–20 unrelated fields were imaged for each animal.

Images were collected with a Leica TCS SP2 AOBS confocal microscope (Mannheim) with  $\times 60$  oil-immersion optics. Laser lines at 458 nm, 488 nm and 633 nm for excitation of CFP, GFP and AF647, respectively, were provided by an Ar laser and a HeNe laser. Detection ranges were set to eliminate crosstalk between fluorophores. Three-dimensional reconstruction was performed using Volocity (Improvision Inc.).

**Statistical analysis.** Statistical analysis methods were the standard two-tailed Student's *t*-test for two data sets and ANOVA followed by Bonferroni/Dunn post

hoc tests for multiple data sets using Prism (GraphPad Inc.), except for human-monocyte transfer with antibody treatment, where a paired *t*-test was used because of variations among different donors. For the spontaneous-metastasis assay of MDA-MB-231 cells, percentage differences in numbers of lung metastases were compared between groups using parametric survival regression methods, with metastasis counts of more than 100 considered censored at 100. *P* values of less than 0.05 were deemed significant.

21. Nielsen, B. S. *et al.* A precise and efficient stereological method for determining murine lung metastasis volumes. *Am. J. Pathol.* **158**, 1997–2003 (2001).
22. Abramoff, M. D., Magelhaes, P. J. & Ram, S. J. Image processing with ImageJ. *Biophotonics Int.* **11**, 36–42 (2004).
23. Havens, A. M. *et al.* An *in vivo* mouse model for human prostate cancer metastasis. *Neoplasia* **10**, 371–380 (2008).
24. Brandt, B. *et al.* 3D-extravasation model — selection of highly motile and metastatic cancer cells. *Semin. Cancer Biol.* **15**, 387–395 (2005).
25. Stern, P. *et al.* A system for Cre-regulated RNA interference *in vivo*. *Proc. Natl Acad. Sci. USA* **105**, 13895–13900 (2008).
26. Im, J. H. *et al.* Coagulation facilitates tumor cell spreading in the pulmonary vasculature during early metastatic colony formation. *Cancer Res.* **64**, 8613–8619 (2004).

# Tumour hypoxia promotes tolerance and angiogenesis via CCL28 and T<sub>reg</sub> cells

Andrea Facciabene<sup>1</sup>, Xiaohui Peng<sup>1</sup>, Ian S. Hagemann<sup>2</sup>, Klara Balint<sup>1</sup>, Andrea Barchetti<sup>1</sup>, Li-Ping Wang<sup>2</sup>, Phyllis A. Gimotty<sup>3</sup>, C. Blake Gilks<sup>4</sup>, Priti Lal<sup>2</sup>, Lin Zhang<sup>1</sup> & George Coukos<sup>1</sup>

**Although immune mechanisms can suppress tumour growth<sup>1,2</sup>, tumours establish potent, overlapping mechanisms that mediate immune evasion<sup>3–6</sup>. Emerging evidence suggests a link between angiogenesis and the tolerance of tumours to immune mechanisms<sup>7–10</sup>. Hypoxia, a condition that is known to drive angiogenesis in tumours, results in the release of damage-associated pattern molecules, which can trigger the rejection of tumours by the immune system<sup>11</sup>. Thus, the counter-activation of tolerance mechanisms at the site of tumour hypoxia would be a crucial condition for maintaining the immunological escape of tumours. However, a direct link between tumour hypoxia and tolerance through the recruitment of regulatory cells has not been established. We proposed that tumour hypoxia induces the expression of chemotactic factors that promote tolerance. Here we show that tumour hypoxia promotes the recruitment of regulatory T (T<sub>reg</sub>) cells through induction of expression of the chemokine CC-chemokine ligand 28 (CCL28), which, in turn, promotes tumour tolerance and angiogenesis. Thus, peripheral immune tolerance and angiogenesis programs are closely connected and cooperate to sustain tumour growth.**

Seventeen human ovarian cancer cell lines were incubated for 16 h in either hypoxic conditions (1.5% O<sub>2</sub>) or oxic conditions (21% O<sub>2</sub>). We used custom quantitative PCR (qPCR) arrays to analyse the changes in expression of chemokines and their receptors, as well as those of other genes implicated in immune regulation (Supplementary Tables 1 and 2). We considered only chemokines that fit two criteria: the chemokines had to be expressed by at least 9 of the 17 tumour lines (at baseline or under hypoxic conditions), and their expression had to show concordant changes under hypoxic conditions in all of the cell lines. CCL28 was the most highly upregulated chemokine gene under hypoxic conditions (Fig. 1a). That CCL28 protein production is regulated by hypoxia and by hypoxia inducible factor 1 $\alpha$  (HIF1 $\alpha$ ) was confirmed in ovarian cancer cell lines *in vitro* (Fig. 1b and Supplementary Fig. 1). *In vivo*, CCL28 expression varied among tumours and was localized mainly to tumour cells (Supplementary Fig. 2). Furthermore, CCL28 was upregulated in areas of tumour hypoxia in tumour xenografts (Supplementary Fig. 3), and CCL28 expression correlated significantly with HIF1 $\alpha$  expression in ovarian cancer samples (Fig. 1c). Similar to HIF1 $\alpha$ <sup>12,13</sup>, CCL28 overexpression was associated with a poor outcome in patients with ovarian cancer (Fig. 1d, e).

CCL28, also known as mucosae-associated epithelial chemokine (MEC), has been implicated in mucosal immunity<sup>14</sup>, and its production is increased by pro-inflammatory cytokines and bacterial products<sup>15</sup>. However, CCL28 also recruits CC-chemokine receptor 10 (CCR10)<sup>+</sup> T<sub>reg</sub> cells during liver inflammation<sup>16</sup>. We examined whether hypoxic tumour cells recruit human T<sub>reg</sub> cells *in vitro* through CCL28. In chemotaxis assays in which freshly isolated human peripheral blood mononuclear cells (PBMCs) were allowed to migrate towards supernatants from tumour cells (Supplementary Fig. 4a), hypoxic medium

recruited significantly more CD4<sup>+</sup>CD25<sup>+</sup>forkhead box P3 (FOXP3)<sup>+</sup> cells than oxic medium (Fig. 2a and Supplementary Fig. 4b). The ability of hypoxic medium to recruit preferentially CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> cells was abrogated by antibody that neutralized human CCL28 (Fig. 2b and Supplementary Fig. 5a).

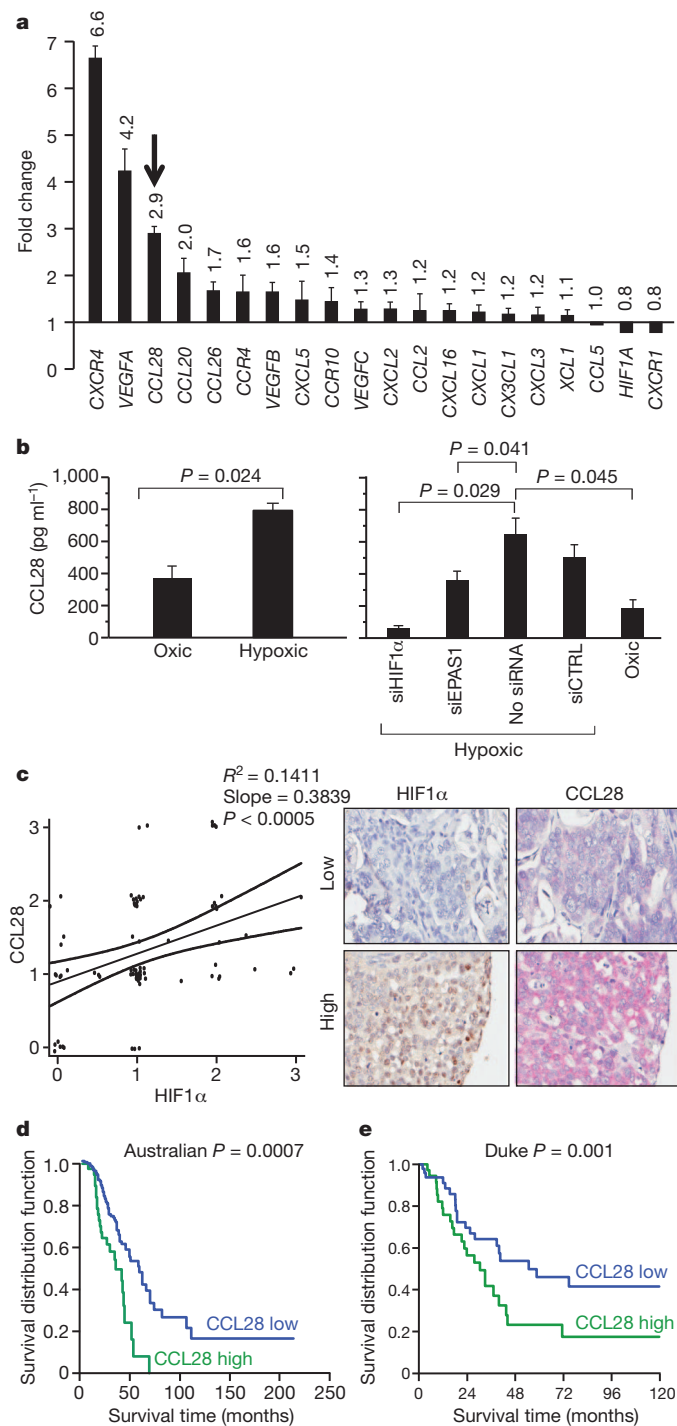
CCR3 and CCR10 are the known receptors for CCL28 (refs 14, 17). The addition of antibody specific for human CCR10 reduced the preferential recruitment of CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> cells by hypoxic tumour cell medium (Fig. 2b and Supplementary Fig. 5a) but did not affect the modest migration induced by oxic tumour cell medium (Supplementary Fig. 5b). Recombinant human CCL28 also preferentially recruited CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> cells from human PBMCs, and this response was abrogated by anti-CCL28 antibody and significantly attenuated by anti-CCR10 antibody but not by a control antibody (Fig. 2c). CCR3 neutralization did not consistently reproduce the same effects as CCR10 neutralization in these experiments. Thus, hypoxic tumour cells recruit T<sub>reg</sub> cells through CCL28, mostly through its binding to CCR10. Consistent with a population containing a higher frequency of T<sub>reg</sub> cells, T cells that were recruited by recombinant human CCL28 showed a significantly dampened response to alloantigen in mixed allogeneic leukocyte reaction assays, which did not occur when PBMCs were pre-incubated with anti-CCR10 antibody (Supplementary Fig. 5c). These results establish a new direct link between tumour cell hypoxia, CCL28 upregulation and human T<sub>reg</sub>-cell recruitment through CCR10.

Human CCL28 is highly homologous to its mouse counterpart<sup>18</sup>. Similar to human ovarian cancer cells, ID8 cells, which are a line of mouse ovarian cancer cells<sup>19,20</sup>, upregulated CCL28 under hypoxic conditions (Supplementary Fig. 6), and CCL28 expression correlated positively with expression of the hypoxia marker carbonic anhydrase IX in orthotopic, intraperitoneal, ID8 tumours (Supplementary Fig. 7). To learn more about the role of CCL28 overexpression, we transduced ID8 cells with mouse Ccl28 (denoted ID8-ccl28) (Supplementary Fig. 6). There were significantly higher levels of CCL28 protein in intraperitoneal ID8-ccl28 tumours (Fig. 3a) and in the peritoneal fluid of these mice (known as ascites) (Fig. 3b) than in control, mock-transduced, ID8 tumours and peritoneal fluid, mimicking human ovarian cancer with a high and low level of CCL28 expression, respectively. ID8-ccl28 tumours accumulated significantly more CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> cells *in vivo* than did ID8 tumours (Fig. 3c). This was a result of direct recruitment, as ascites from mice with ID8-ccl28 tumours recruited significantly more CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> cells from mixed splenocytes *in vitro* than did ascites from mice with ID8 tumours (Fig. 3d). Supporting a role for CCL28 and CCR10, ~90% of T<sub>reg</sub> cells in ascites from mice with ID8-ccl28 tumours were CCR10<sup>+</sup> (Supplementary Fig. 8).

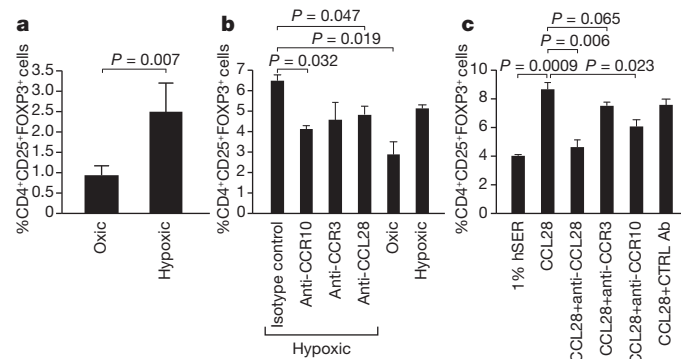
Orthotopic, intraperitoneal, ID8-ccl28 tumours showed significantly faster growth and induced faster ascites development than ID8 tumours (Fig. 3e). To test which subset of CCL28-recruited cells

<sup>1</sup>Ovarian Cancer Research Center, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. <sup>2</sup>Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. <sup>3</sup>Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. <sup>4</sup>Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, V5Z 1M9, Canada.





**Figure 1 | CCL28 in tumours is upregulated by hypoxia.** **a**, Gene expression changes in seven primary and ten established ovarian cancer cell lines following 16 h in hypoxic conditions (low-density qPCR array data, normalized to 18S ribosomal RNA). **b**, CCL28 protein in supernatants from SKOV3 ovarian cancer cells incubated under hypoxic or oxic conditions, as determined by ELISA (left). Cells were transfected with short interfering (siRNA) directed against HIF1 $\alpha$  (siHIF1 $\alpha$ ) or HIF2 $\alpha$  (also known as EPAS1) (siEPAS1) or with control scrambled siRNA (siCTRL) 24 h before being subjected to hypoxia (right). **c**, Correlation of CCL28 and HIF1 $\alpha$  protein expression in ovarian cancers, as determined by using immunohistochemistry (left; best-fit line is indicated, together with 95% confidence bands). Representative images from tumours expressing large (high) or small (low) amounts of CCL28 (red) and HIF1 $\alpha$  (brown) (right). **d**, **e**, CCL28 overexpression is associated with shorter survival in patients with ovarian cancer: GSE9891 data,  $n = 220$  (**d**); and GSE3149 data,  $n = 133$  (**e**). **a**, **b**, Error bars, s.e.m.

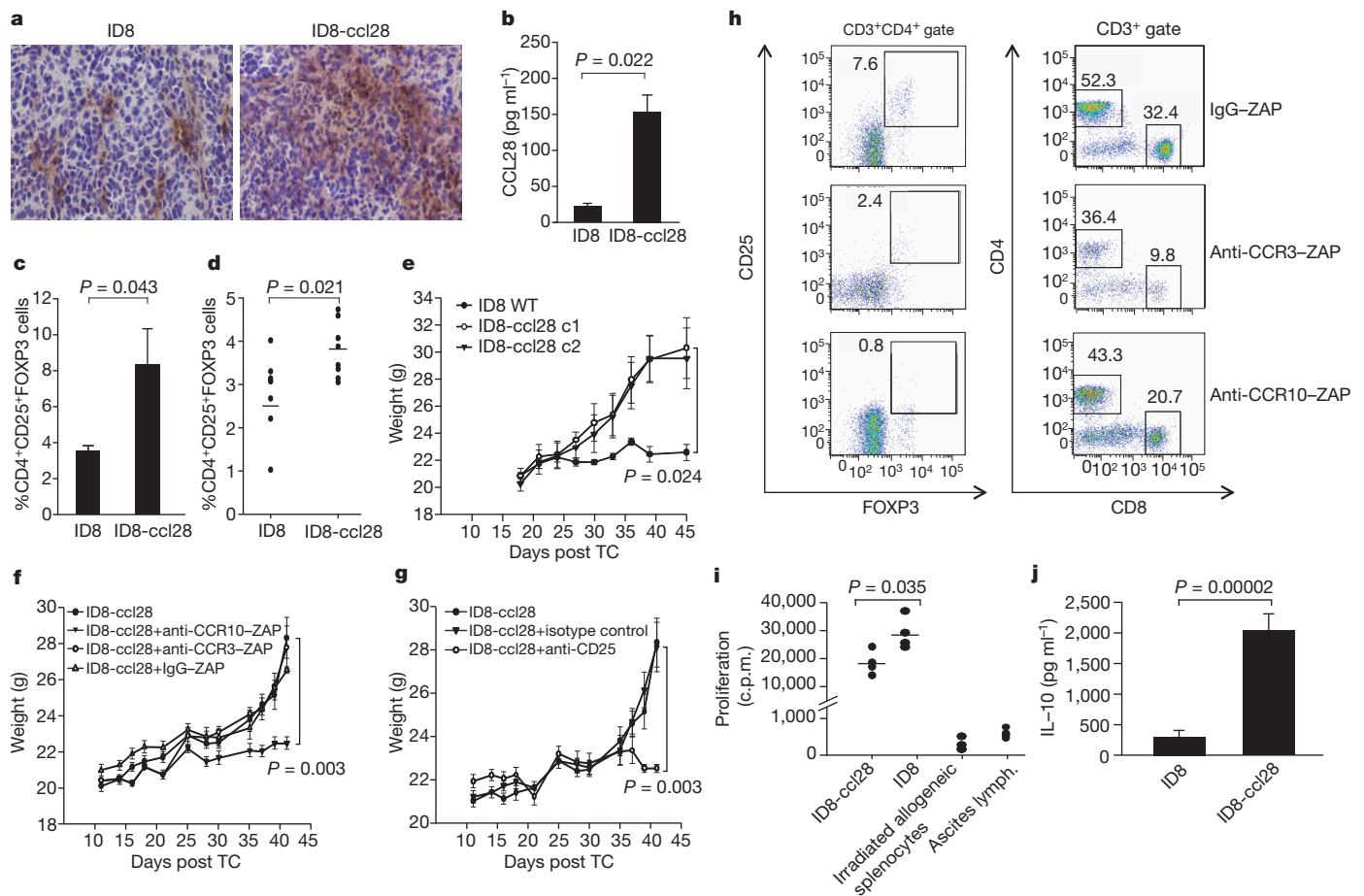


**Figure 2 | Hypoxic tumour cells recruit CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> T<sub>reg</sub> cells through CCL28-CCR10.** **a**, Hypoxic OVCAR5 cell supernatants recruit more T<sub>reg</sub> cells than oxic cell supernatants. **b**, T<sub>reg</sub>-cell migration towards hypoxic medium is attenuated by blockade of CCL28 or CCR10. **c**, Human recombinant CCL28 (denoted CCL28 in the figure) recruits T<sub>reg</sub> cells, and this recruitment is abrogated by blockade of CCL28 or CCR10. CTRLAb, control antibody; hSER, 1% human serum control medium. In all panels, the y axis represents the percentage of CD25<sup>+</sup>FOXP3<sup>+</sup> T<sub>reg</sub> cells among the CD4<sup>+</sup> cells recruited to the lower chamber of chemotaxis chambers from human PBMCs seeded in the upper chambers, as determined by flow cytometry gating on CD3<sup>+</sup>CD4<sup>+</sup>7-AAD<sup>-</sup> cells. Error bars, s.e.m.

was responsible for accelerating the growth of ID8-ccl28 tumours, we depleted CCR3<sup>+</sup> cells or CCR10<sup>+</sup> cells by using saporin (ZAP)-conjugated anti-mouse CCR3 antibody (denoted anti-CCR3-ZAP) or anti-CCR10-ZAP (Fig. 3f). A single intraperitoneal injection of the ZAP-conjugated antibody (40  $\mu$ g) depleted >90% of CCR10<sup>+</sup> or CCR3<sup>+</sup> cells within 72 h (Supplementary Fig. 9). Mice then received the anti-CCR10-ZAP or anti-CCR3-ZAP immunotoxin 2 days before and 8 days after intraperitoneal inoculation with ID8-ccl28 tumours. Anti-CCR10-ZAP suppressed tumour growth and abrogated the effects of CCL28 overexpression, whereas anti-CCR3-ZAP had no effect on tumour growth (Fig. 3f). Importantly, ID8-ccl28 cells expressed no CCR10 or CCR3, and antibody-ZAP conjugates had no direct effect on tumour growth *in vitro* (Supplementary Fig. 10). Both anti-CCR10-ZAP and anti-CCR3-ZAP effectively eliminated systemic CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> T<sub>reg</sub> cells within 72 h of injection. Importantly, however, anti-CCR10-ZAP had a relatively minor effect on CD8<sup>+</sup> cells, thus increasing the CD8<sup>+</sup> cell to T<sub>reg</sub>-cell ratio up to fivefold compared with IgG-ZAP (Fig. 3h). By contrast, consistent with a lack of antitumour efficacy, anti-CCR3-ZAP depleted T<sub>reg</sub> cells but also a large proportion of CD8<sup>+</sup> cells, maintaining a constant CD8<sup>+</sup> cell to T<sub>reg</sub>-cell ratio (Fig. 3h). This could be explained by the expression of cognate receptors on T-cell subsets; a significant proportion of CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> T<sub>reg</sub> cells are CCR3<sup>+</sup> and/or CCR10<sup>+</sup>, whereas CD8<sup>+</sup> cells express CCR3 (>75% positive) but not CCR10 (<3% positive) (Supplementary Fig. 11). Thus, specific depletion of CCR10<sup>+</sup> cells abrogated the tumour-promoting effects of CCL28.

Because leukocytes other than T<sub>reg</sub> cells could be recruited by CCL28, we tested the contribution of CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> T<sub>reg</sub> cells to the rapid growth of ID8-ccl28 tumours by using anti-CD25 antibody<sup>21</sup>, which depleted most of the CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> T<sub>reg</sub> cells (Supplementary Fig. 12). CD25<sup>+</sup> T-cell depletion hindered tumour growth and abrogated the effects of CCL28 overexpression (Fig. 3g). Importantly, ID8 or ID8-ccl28 cells did not express CD25, and the anti-CD25 antibody had no direct effect on their growth *in vitro* (Supplementary Fig. 10), indicating an extrinsic effect *in vivo* (mediated through T<sub>reg</sub>-cell depletion). Thus, a direct link exists between tumour CCL28 upregulation and accelerated tumour growth, which is specifically attributable to T<sub>reg</sub>-cell recruitment *in vivo* through CCR10.

In line with the current understanding of T<sub>reg</sub>-cell function and with our *in vitro* results showing that human lymphocytes recruited by recombinant human CCL28 showed dampened reactivity (Fig. 2c



**Figure 3 | CCL28 promotes tumour growth through attracting CCR10<sup>+</sup> T<sub>reg</sub> cells.** **a**, CCL28 (brown) in intraperitoneal ID8 and ID8-ccl28 tumours, as determined by immunohistochemistry (4',6-diamidino-2-phenylindole (DAPI), blue). **b**, CCL28 in ID8 and ID8-ccl28 tumour ascites, as determined by ELISA. **c**, CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> cells in ID8 or ID8-ccl28 ascites. **d**, Spleen CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> cells recruited *in vitro* by ID8 or ID8-ccl28 tumours (two clones, c1 and c2) ( $n = 40$  per group). Weight is a reliable marker of tumour growth. TC, tumour challenge. **e**, Weights of mice bearing ID8 or ID8-ccl28 tumours that were untreated or treated with anti-CCR3-ZAP antibody, anti-CCR10-ZAP antibody or control IgG-ZAP ( $n = 10$  per group). **f**, Weights of mice bearing ID8-ccl28 tumours that were untreated or treated with anti-CCR3-ZAP antibody, anti-CCR10-ZAP antibody or control IgG-ZAP ( $n = 10$  per group). **g**, Weights of mice bearing ID8-ccl28 tumours

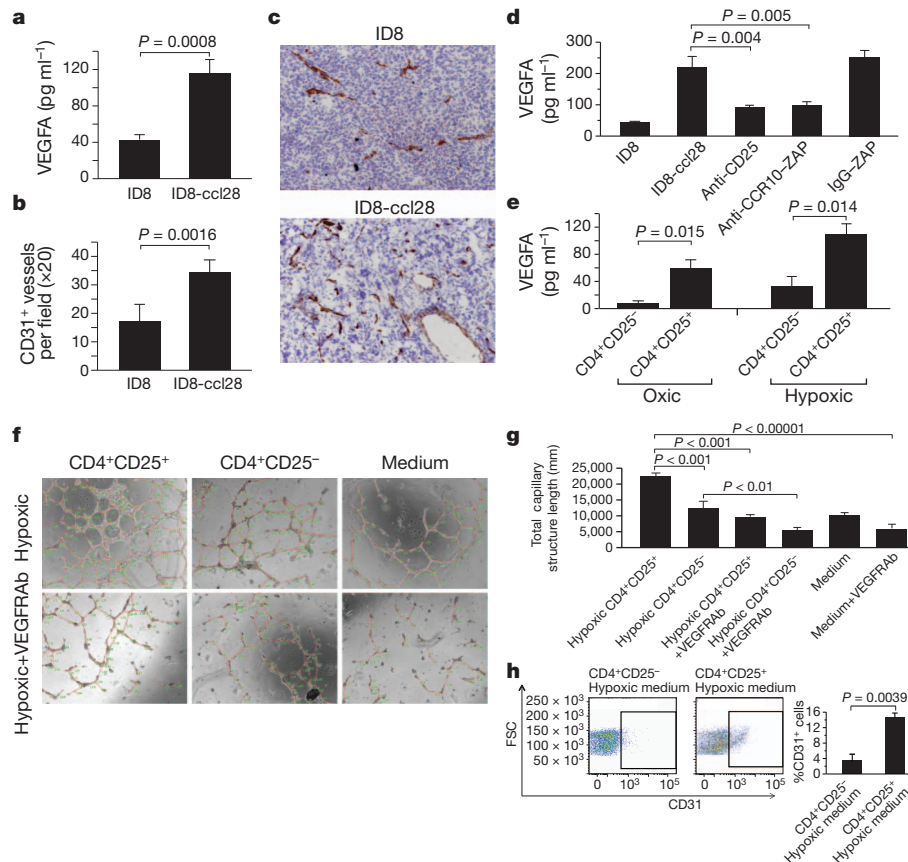
and Supplementary Fig. 5c), responder T cells isolated from the ascites of mice with ID8-ccl28 tumours showed less proliferation than T cells isolated from the ascites of mice with ID8 tumours in response to irradiated allogeneic splenocyte targets (Fig. 3i). Thus, the tumour-derived CCL28 recruits more CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> T<sub>reg</sub> cells, which suppress effector T cell function. Consistent with a more tolerogenic environment, we observed markedly higher interleukin-10 (IL-10) levels (Fig. 3j) in the ascites of mice with ID8-ccl28 tumours than in those with control ID8 tumours.

We have reported an inverse correlation between angiogenesis and tumour-infiltrating T cells<sup>2,7,22,23</sup>. We also found significantly increased amounts of vascular endothelial growth factor A (VEGFA) in the ascites of mice with ID8-ccl28 tumours than those with ID8 tumours (Fig. 4a), and intraperitoneal ID8-ccl28 solid tumour nodules showed significantly increased microvascular density relative to control ID8 tumours (Fig. 4b, c). ID8 and ID8-ccl28 cells express similar amounts of VEGFA (Supplementary Fig. 13). Confirming that excess VEGFA was contributed by CCR10<sup>+</sup> haematopoietic cells recruited by CCL28, there was a significant reduction in tumour VEGFA levels (Fig. 4d) and a significant reduction in tumour microvascular density (Supplementary Fig. 14) in mice that received anti-CCR10-ZAP relative to mice that received control immunotoxin (IgG-ZAP). Further supporting

that were untreated or treated with anti-CD25 antibody or IgG isotype control ( $n = 10$  per group). **h**, CCR10 depletion eliminates most CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> cells but not CD8<sup>+</sup> T cells, as determined by flow cytometric analysis. CCR3 depletion eliminates both populations (numbers inside plots refer to the boxed areas: left column, % T<sub>reg</sub> cells; right column, % CD3<sup>+</sup>CD4<sup>+</sup> cells (left) and % CD3<sup>+</sup>CD8<sup>+</sup> cells (right)). **i**, Responder T cells from the ascites of ID8 tumours proliferate more than T cells from the ascites of ID8-ccl28 tumours in a mixed leukocyte reaction. c.p.m., counts per minute of incorporated [<sup>3</sup>H]thymidine. Ascites lymph., ascites lymphocytes. **j**, IL-10 in ID8 and ID8-ccl28 ascites, as determined by ELISA ( $n = 9$  per group). **b–g**, **i**, **j**, Error bars, s.e.m.

the role of T<sub>reg</sub> cells, there was a significant reduction in tumour VEGFA levels (Fig. 4d) and a significant reduction in tumour microvascular density (Supplementary Fig. 14) in mice that received anti-CD25 antibody relative to mice that received control IgG. Thus, T<sub>reg</sub>-cell recruitment has a key role in establishing a VEGFA-rich tumour microenvironment and increasing tumour angiogenesis, whereas the depletion of T<sub>reg</sub> cells reduces tumour VEGFA levels and tumour vascularization.

We found that T<sub>reg</sub> cells can also directly contribute to the VEGFA pool in the tumour microenvironment. CD4<sup>+</sup>CD25<sup>+</sup> cells purified from fresh human donor PBMCs secreted markedly more VEGFA than CD4<sup>+</sup>CD25<sup>-</sup> cells under oxic or hypoxic conditions (Fig. 4e). Similar results were obtained with mouse T<sub>reg</sub> cells purified from the spleen (data not shown). Furthermore, medium conditioned by hypoxic human peripheral blood CD4<sup>+</sup>CD25<sup>+</sup> cells induced a significantly larger expansion of human umbilical vein endothelial cells, as assessed by the total length of capillary endothelial networks formed *in vitro*, than medium conditioned by hypoxic human CD4<sup>+</sup>CD25<sup>-</sup> cells (Fig. 4f, g). This effect was mediated by VEGFA, because it was abrogated by neutralizing antibodies against human VEGF receptors 1 and 2 (VEGFRab, Fig. 4f, g). Last, medium conditioned by purified hypoxic mouse spleen CD4<sup>+</sup>CD25<sup>+</sup> or CD4<sup>+</sup>CD25<sup>-</sup> cells promoted



**Figure 4** | **T<sub>reg</sub> cells promote tumour angiogenesis.** **a**, VEGFA protein in the ascites from intraperitoneal ID8 or ID8-ccl28 tumours, as determined by ELISA. **b**, Vasculature density in ID8 and ID8-ccl28 tumours, as determined by immunohistochemistry (×20, original magnification). **c**, Vasculature in representative ID8 and ID8-ccl28 tumours, using CD31 (brown) immunohistochemistry (DAPI, blue). **d**, Less VEGFA protein is present in the ascites of ID8-ccl28 tumours following administration of anti-CD25 antibody or anti-CCR10-ZAP, as determined by ELISA. IgG-ZAP, control immunotoxin. **e**, VEGFA expression by human CD4<sup>+</sup>CD25<sup>+</sup> or CD4<sup>+</sup>CD25<sup>-</sup>

angiogenesis *in vivo*. Cell-free, growth-factor-free Matrigel plugs enriched with supernatants from hypoxic mouse CD4<sup>+</sup>CD25<sup>+</sup> splenocytes accumulated significantly more CD31<sup>+</sup> endothelial cells over 3 days (~15% of total accumulated cells on average) than Matrigel enriched with hypoxic mouse CD4<sup>+</sup>CD25<sup>-</sup> splenocyte supernatants (~3.8% of total cells accumulated) (Fig. 4h). Thus, T<sub>reg</sub> cells constitutively secrete VEGFA, which is further upregulated by hypoxia, and promote a pro-angiogenic tumour milieu.

Here we provide the first demonstration that hypoxic intraperitoneal tumours recruit CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> T<sub>reg</sub> cells, which dampen effector T cell function and promote tumour angiogenesis through VEGFA. This finding reinforces the link between tumour hypoxia, peripheral tolerance and angiogenesis. In addition to adenosine<sup>24</sup>, hypoxic tumour cells promote tolerance by secreting CCL28 and recruiting T<sub>reg</sub> cells to hypoxic areas. There, the T<sub>reg</sub>-cell suppressive function can be enhanced<sup>25</sup>, while the T<sub>reg</sub> cells promote angiogenesis. Importantly, VEGFA can further promote tumour tolerance<sup>8–10,26</sup>. Although T<sub>reg</sub> cells can contribute directly to excess production of VEGFA and can support endothelial cell recruitment and expansion, other tolerogenic leukocyte populations such as myeloid-derived suppressor cells<sup>27,28</sup> and plasmacytoid dendritic cells<sup>28,29</sup> also produce VEGFA and support tumour angiogenesis. However, CD25<sup>+</sup> T<sub>reg</sub> cells are a crucial population, as their elimination abrogated VEGFA overexpression in ovarian tumours. Thus, the tumour immune tolerance and angiogenesis programs are closely connected at many levels and work hand in hand to ensure tumour growth.

T cells after 16 h under hypoxic conditions. **f**, **g**, Endothelial-tube formation by human umbilical vein endothelial cells incubated in supernatants from hypoxic human CD4<sup>+</sup>CD25<sup>+</sup> T cells, hypoxic human CD4<sup>+</sup>CD25<sup>-</sup> T cells or medium, all with or without VEGFR1/VEGFR2-neutralizing antibodies (VEGFRAb). **h**, CD31<sup>+</sup> endothelial cells in 72-h subcutaneous Matrigel plugs enriched with hypoxic mouse CD4<sup>+</sup>CD25<sup>+</sup> or CD4<sup>+</sup>CD25<sup>-</sup> T-cell medium (*n* = 5 per group), as determined by flow cytometric analysis gating on 7-AAD<sup>-</sup>CD45<sup>-</sup> cells. FSC, forward scatter. Right, % CD31<sup>+</sup> cells is the percentage of CD31<sup>+</sup> cells among the 7-AAD<sup>-</sup>CD45<sup>-</sup> cells. **a**, **b**, **d**, **e**, **g**, **h**, Error bars, s.e.m.

## METHODS SUMMARY

We used early-passage primary cell lines from four solid ovarian cancers and three ascites<sup>30</sup> and ten established human ovarian cancer cell lines. ID8-ccl28 cells were derived from ID8 cells that had been transfected with codon-optimized *Ccl28* cDNA cloned into the vector pcDNA3. For hypoxia experiments, cells were cultured for 16 h under hypoxic conditions (1.5% O<sub>2</sub>) or oxic conditions (21% O<sub>2</sub>), both with 5% CO<sub>2</sub> at 37 °C. For tissue studies, we used a tissue microarray comprising 88 advanced-stage ovarian cancer samples. We used 6–8-week-old female C57BL/6 mice to establish intraperitoneal ID8 or ID8-ccl28 tumours. *In vivo* depletion of CD4<sup>+</sup>CD25<sup>+</sup> cells was achieved by intraperitoneal administration of anti-CD25 antibody or an immunotoxin consisting of anti-mouse CCR10 or anti-mouse CCR3 antibody conjugated at an equimolar ratio to streptavidin-ZAP. Experiments were performed at least three times with ten animals per group. For detecting CCL28 in EF5<sup>+</sup> (2-(2-nitro-1H-imidazol-1-yl)-N-(2,2,3,3,3-pentafluoropropyl) acetamide-positive) hypoxic areas in human ovarian tumour xenografts *in vivo*, we used 8-week-old NOD.Cg-Prkdc<sup>scid</sup> Il2rg<sup>tm1Wjl</sup>/SzJ (NSG) mice. TaqMan Low Density Arrays (384 wells) were custom designed to comprise 190 genes involved in immune regulation. The Methods describes the protocols for the following in detail: western blotting, enzyme-linked immunosorbent assays (ELISAs), protein array analysis, tissue immunostaining, flow cytometric analysis, migration assays, proliferation assays, mixed lymphocyte reactions, endothelial-tube formation assays and *in vivo* angiogenesis assessments. We performed pairwise comparisons using Student's *t*-test for independent groups. We used Spearman's correlations and linear regression to estimate the correlation between immunohistochemistry parameters. Two publicly available Affymetrix array expression data sets (GSE3149 and GSE9891), covering 353 human ovarian cancer patients<sup>31,32</sup>, were mined to analyse the correlation between *CCL28* and survival. An optimal cut-off point for *CCL28* gene expression defining two groups of



patients with different survival curves was determined using the program X-tile<sup>33</sup>. The log-rank test was used to determine whether the survival curves were significantly different.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 18 April 2010; accepted 3 May 2011.**

- Koebe, C. M. *et al.* Adaptive immunity maintains occult cancer in an equilibrium state. *Nature* **450**, 903–907 (2007).
- Zhang, L. *et al.* Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. *N. Engl. J. Med.* **348**, 203–213 (2003).
- Zou, W. Immunosuppressive networks in the tumour environment and their therapeutic relevance. *Nature Rev. Cancer* **5**, 263–274 (2005).
- Curiel, T. J. *et al.* Specific recruitment of regulatory T cells in ovarian carcinoma fosters immune privilege and predicts reduced survival. *Nature Med.* **10**, 942–949 (2004).
- Nagaraj, S. & Gabrilovich, D. I. Myeloid-derived suppressor cells. *Adv. Exp. Med. Biol.* **601**, 213–223 (2007).
- Marigo, I., Dolcetti, L., Serafini, P., Zanovello, P. & Bronte, V. Tumor-induced tolerance and immune suppression by myeloid derived suppressor cells. *Immunol. Rev.* **222**, 162–179 (2008).
- Buckanovich, R. J. *et al.* Endothelin B receptor mediates the endothelial barrier to T cell homing to tumors and disables immune therapy. *Nature Med.* **14**, 28–36 (2008).
- Gabrilovich, D. I. *et al.* Production of vascular endothelial growth factor by human tumors inhibits the functional maturation of dendritic cells. *Nature Med.* **2**, 1096–1103 (1996).
- Li, B. *et al.* Vascular endothelial growth factor blockade reduces intratumoral regulatory T cells and enhances the efficacy of a GM-CSF-secreting cancer immunotherapy. *Clin. Cancer Res.* **12**, 6808–6816 (2006).
- Osada, T. *et al.* The effect of anti-VEGF therapy on immature myeloid cell and dendritic cells in cancer patients. *Cancer Immunol. Immunother.* **57**, 1115–1124 (2008).
- Lotfi, R., Lee, J. J. & Lotze, M. T. Eosinophilic granulocytes and damage-associated molecular pattern molecules (DAMPs): role in the inflammatory response within tumors. *J. Immunother.* **30**, 16–28 (2007).
- Shimogai, R. *et al.* Expression of hypoxia-inducible factor 1 $\alpha$  gene affects the outcome in patients with ovarian cancer. *Int. J. Gynecol. Cancer* **18**, 499–505 (2008).
- Daponte, A. *et al.* Prognostic significance of hypoxia-inducible factor 1 $\alpha$  (HIF-1 $\alpha$ ) expression in serous ovarian cancer: an immunohistochemical study. *BMC Cancer* **8**, 335 (2008).
- Pan, J. *et al.* A novel chemokine ligand for CCR10 and CCR3 expressed by epithelial cells in mucosal tissues. *J. Immunol.* **165**, 2943–2949 (2000).
- Feng, N. *et al.* Redundant role of chemokines CCL25/TECK and CCL28/MEC in IgA<sup>+</sup> plasmablast recruitment to the intestinal lamina propria after rotavirus infection. *J. Immunol.* **176**, 5749–5759 (2006).
- Eksteen, B. *et al.* Epithelial inflammation is associated with CCL28 production and the recruitment of regulatory T cells expressing CCR10. *J. Immunol.* **177**, 593–603 (2006).
- Wang, W. *et al.* Identification of a novel chemokine (CCL28), which binds CCR10 (GPR2). *J. Biol. Chem.* **275**, 22313–22323 (2000).
- Lazarus, N. H. *et al.* A common mucosal chemokine (mucosae-associated epithelial chemokine/CCL28) selectively attracts IgA plasmablasts. *J. Immunol.* **170**, 3799–3805 (2003).
- Roby, K. F. *et al.* Development of a syngeneic mouse model for events related to ovarian cancer. *Carcinogenesis* **21**, 585–591 (2000).
- Zhang, L. *et al.* Generation of a syngeneic mouse model to study the effects of vascular endothelial growth factor in ovarian carcinoma. *Am. J. Pathol.* **161**, 2295–2309 (2002).
- Elia, L. *et al.* CD4<sup>+</sup>CD25<sup>+</sup> regulatory T-cell-inactivation in combination with adenovirus vaccines enhances T-cell responses and protects mice from tumor challenge. *Cancer Gene Ther.* **14**, 201–210 (2007).
- Kandalaf, L. E., Facciabene, A., Buckanovich, R. J. & Coukos, G. Endothelin B receptor, a new target in cancer immune therapy. *Clin. Cancer Res.* **15**, 4521–4528 (2009).
- Buckanovich, R. J. *et al.* Tumor vascular proteins as biomarkers in ovarian cancer. *J. Clin. Oncol.* **25**, 852–861 (2007).
- Ohta, A. *et al.* A2A adenosine receptor protects tumors from antitumor T cells. *Proc. Natl Acad. Sci. USA* **103**, 13132–13137 (2006).
- Ben-Shoshan, J., Maysel-Auslender, S., Mor, A., Keren, G. & George, J. Hypoxia controls CD4<sup>+</sup>CD25<sup>+</sup> regulatory T-cell homeostasis via hypoxia-inducible factor-1 $\alpha$ . *Eur. J. Immunol.* **38**, 2412–2418 (2008).
- Curiel, T. J. *et al.* Blockade of B7-H1 improves myeloid dendritic cell-mediated antitumor immunity. *Nature Med.* **9**, 562–567 (2003).
- Yang, L. *et al.* Expansion of myeloid immune suppressor Gr<sup>+</sup>CD11b<sup>+</sup> cells in tumor-bearing host directly promotes tumor angiogenesis. *Cancer Cell* **6**, 409–421 (2004).
- Shojaei, F. *et al.* Tumor refractoriness to anti-VEGF treatment is mediated by CD11b<sup>+</sup>Gr<sup>+</sup> myeloid cells. *Nature Biotechnol.* **25**, 911–920 (2007).
- Curiel, T. J. *et al.* Dendritic cell subsets differentially regulate angiogenesis in human ovarian cancer. *Cancer Res.* **64**, 5535–5538 (2004).
- Bertozzi, C. C. *et al.* Multiple initial culture conditions enhance the establishment of cell lines from primary ovarian cancer specimens. *In Vitro Cell. Dev. Biol. Anim.* **42**, 58–62 (2006).
- Bild, A. H. *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353–357 (2006).
- Tothill, R. W. *et al.* Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin. Cancer Res.* **14**, 5198–5208 (2008).
- Camp, R. L., Dolled-Filhart, M. & Rimm, D. L. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin. Cancer Res.* **10**, 7252–7259 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by National Institutes of Health grant R01-CA116779; National Cancer Institute Ovarian SPORE grant P01-CA83638; and the Ovarian Cancer Research Fund. We thank M. Celeste Simon and S. Evans for generous help with the hypoxia studies; M. Feldman and the University of Pennsylvania Tumor Tissue & Biospecimen Bank for tumour processing; and G. Danet-Desnoyers and the University of Pennsylvania Xenograft Core Facility for NSG mice.

**Author Contributions** A.F. designed many of the experiments and conducted most of them, analysed the data and drafted the manuscript. X.P. assisted with the experiments. I.S.H. analysed and interpreted the tissue-based studies. A.B. conducted the initial hypoxia and qPCR array experiments. K.B. assisted with the experiments. L.-P.W. carried out the tissue stains. P.A.G. analysed the Affymetrix data. C.B.G. assisted with the tissue-based studies. P.L. prepared the tissue microarrays. L.Z. assisted with the study design and provided many cell lines. G.C. conceived and supervised the study, and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to G.C. ([gcks@mail.med.upenn.edu](mailto:gcks@mail.med.upenn.edu)).

## METHODS

**Cell cultures.** Early-passage primary ovarian cancer cell lines from solid tumours (OV43, OV682, OV684 and OV79) or ascites (OV614, OV62 and OV77) were provided by R. G. Carroll and had been developed at the University of Pennsylvania from chemotherapy-naïve stage III or IV ovarian cancer samples, as previously reported<sup>30</sup>. The above-listed primary ovarian cancer cell lines, and the established human ovarian cancer cell lines A1847, A2008, A2780, C200, C70, CP30, OAW42, PE01, PEO4, OVCAR5, SKOV3 and UPN251, as well as ID8 and ID8-ccl28 cells, were propagated in 5% CO<sub>2</sub> at 37 °C in DMEM supplemented with 10% FBS (HyClone, lot APD21174), 100 U ml<sup>-1</sup> penicillin and 100 mg ml<sup>-1</sup> streptomycin. For hypoxia experiments, cells were seeded into 6-well plates at 60% confluence, incubated overnight and then placed into a Heracell 240 incubator (Thermo Scientific) for 16 h under hypoxic (1.5% O<sub>2</sub>) or oxic (21% O<sub>2</sub>) conditions, both with 5% CO<sub>2</sub> at 37 °C. For the analyses using PCR arrays, we used the following 17 cell lines: OV43, OV682, OV684, OV79, OV614, OV62, OV77, A1847, A2008, A2780, C200, C70, CP30, OAW42, PE01, PEO4, SKOV3 and UPN251. Validation hypoxia experiments were repeated with PEO4, SKOV3 and OVCAR5 cell lines. All *in vitro* validation experiments were conducted at least twice in triplicate.

In some experiments, ID8 and ID8-ccl28 cells were incubated with a monoclonal anti-CD25 antibody (PC61, 1 µg ml<sup>-1</sup>), which was purified using a protein G column (Amersham) from a PC61 hybridoma (ATCC) developed in nude mice. Alternatively, ID8 and ID8-ccl28 cells were incubated with antibody specific for mouse CCR10 or CCR3 (1 µg ml<sup>-1</sup>; anti-mouse CCR10, clone 248918; anti-mouse CCR3, clone 61828; R&D Systems) that had been conjugated at an equimolar ratio to streptavidin-ZAP (Advanced Targeting Systems). The cells were then washed and cultured for up to 9 days. At the end of the culture time, cell numbers were assessed by Trypan blue exclusion.

For hypoxia experiments using T cells, mouse spleen-derived CD4<sup>+</sup> cells were cultured in RPMI-1640 medium (Gibco) containing 10% FBS (HyClone), and human peripheral-blood-derived CD4<sup>+</sup> cells were cultured in AIM V medium (Gibco) containing 5% human AB serum (Valley Biochemical) using hypoxic conditions identical to those above. All *in vitro* experiments were conducted at least twice in triplicate.

**Human tumour samples.** We conducted tissue-based analyses using fresh specimens of stage III or IV epithelial ovarian cancer that differed from the samples used to develop the above primary cell lines. Tumour tissues were snap frozen in liquid nitrogen and stored at -80 °C until use for western blotting analyses. A tissue microarray was developed at the University of Pennsylvania Tissue Microarray Facility of the Department of Pathology, by using a series of 88 tumour samples from 53 treatment-naïve patients with stage IIIC or IV papillary serous epithelial ovarian cancer who underwent primary resection at our institute between 2005 and 2008. Slides stained with haematoxylin and eosin were reviewed and annotated by a trained pathologist, and paraffin-embedded tissue blocks were selected to construct a tissue microarray. For each block, triplicate 0.6-mm cores of tumour were placed on a tissue microarray using a manual arrayer. This tissue microarray was used for CCL28, pan-cytokeratin and HIF1α immunostaining. All specimens were processed in compliance with the institutional review board and the US Health Insurance Portability and Accountability Act (HIPAA) requirements.

**CCL28 cloning and transfection.** Mouse *Ccl28* cDNA (GENEART) was cloned into the pcDNA3 expression vector (Invitrogen). Five micrograms of pcDNA3-ccl28 in 95 µl Opti-MEM (Invitrogen) was mixed with 3 µl Lipofectamine 2000 (Invitrogen) in 97 µl Opti-MEM and incubated at room temperature for 30 min. The mixture was added to 90% confluent ID8 cells for 6 h at 37 °C. Following transfection, cells were seeded into 96-well plates at different concentrations, ranging from 5 to 200 cells per well, to generate several multiclonal populations of ID8-ccl28-transduced cell lines. We repeated all *in vivo* experiments with two different, randomly selected lines, c1 and c2, which had been developed from 5 and 50 initial ID8 cells, respectively. The growth of these two ID8-ccl28 lines was identical *in vivo* (Supplementary Fig. 12).

**Mouse studies.** Six-eight-week-old female C57BL/6 mice (Charles River) were injected intraperitoneally with 5 × 10<sup>6</sup> ID8 or ID8-ccl28 cells. After the appearance of ascites, animals were weighed twice a week, as weight is a reliable measure of tumour growth in this model. Ascites were collected by paracentesis and used for cellular and molecular analyses when animals in each group reached a weight of ~30 g. Solid tumours were also collected for analysis when animals reached a weight of ~30 g in each group. *In vivo* depletion of CD4<sup>+</sup>CD25<sup>+</sup> cells was achieved by intraperitoneal injection of a monoclonal anti-CD25 antibody (PC61, 400 µg per mouse), which had been purified using a protein G column (Amersham) from a PC61 hybridoma (ATCC) that was developed in nude mice. The efficiency of CD4<sup>+</sup>CD25<sup>+</sup> cell depletion was assessed by using spleen cell analyses. CCR10<sup>+</sup> or CCR3<sup>+</sup> cells were depleted by intraperitoneal injection of an immunotoxin that was constructed with anti-mouse CCR10 or anti-mouse CCR3

antibody, respectively (40 µg per mouse), as described above. The efficiency of CCR3<sup>+</sup> and CCR10<sup>+</sup> cell depletion was assessed in intraperitoneal fluid and ascitic fluid by flow cytometry. Experiments were performed at least three times, with ten animals per group; if the data from three independent experiments were concordant, the results were considered conclusive and analysed statistically.

For detection of CCL28 in hypoxic areas in human ovarian tumours *in vivo*, 8-week-old NOD.Cg-Prkdc<sup>scid</sup> Il2rg<sup>tm1Wjl</sup>/SzJ (NSG) mice, provided by the Xenograft Core Facility at the University of Pennsylvania, were transplanted subcutaneously with 3 × 10<sup>6</sup> human ovarian cancer (OVCAR5) cells. After the tumours reached 10 mm in diameter, animals were injected intravenously with 250 µl EF5 (100 mM solution). After 170 min, 100 µl Hoechst dye (100 mM, Sigma) was injected intravenously. After 10 min, the animals were killed, and their tumours were excised and immediately embedded in OCT medium. Tumour sections were subjected to double immunofluorescence staining for EF5 and human CCL28 (using an antibody from R&D Systems), as previously described<sup>34</sup>.

**RNA isolation and quantitative RT-PCR.** TaqMan Low Density Arrays (384 well, Applied Biosystems) were custom designed to comprise 190 genes involved in immune regulation, including those encoding cytokines, growth factors and chemokines, as well as their receptors, several antimicrobial peptides, co-stimulatory molecules, negative stimulatory molecules of the B7 family and lineage markers. Each of the 17 cell lines detailed above was cultured in triplicate wells under oxic or hypoxic conditions, as described above. Total RNA was immediately isolated from oxic or hypoxic cells at the end of the experiment, using TRIzol reagent (Invitrogen). The RNA concentration was measured with a 2100 Bioanalyzer (Agilent) using an RNA 6000 Nano LabChip. Total RNA (5 µg) was reverse transcribed using High Capacity cDNA Reverse Transcription Kits (Applied Biosystems), according to the manufacturer's instructions. Single-stranded cDNA was generated from each culture well, and cDNAs were pooled for each cell line for each condition. cDNA was combined with 50 µl TaqMan Universal PCR Master Mix and water, and was then loaded on custom-designed, 384-well TaqMan Low Density Arrays in duplicate, followed by loading of 100 µl sample per port. For a complete list of the included genes, see Supplementary Table 1; primer sequences are listed in Supplementary Table 2. Thermal cycling conditions were as follows: 50 °C for 2 min, 95 °C for 10 min, 95 °C for 15 s and 60 °C for 1 min. Samples were analysed using the 7900HT system with TaqMan LDA Upgrade (Applied Biosystems) and SDS software (version 2.2). The expression level of each gene was normalized to 18S rRNA. Each gene was assessed in duplicate in every experiment, and only the genes with reproducible amplification curves were analysed. Duplicate expression levels for each gene were averaged when concordant, and the hypoxia expression level was calibrated against the oxic control sample to obtain the fold change induced by hypoxia. Conventional quantitative PCR with reverse transcription (RT-PCR) was performed as detailed elsewhere<sup>23</sup>. All transcripts were confirmed by electrophoresis in 3% agarose gels. Mouse *Ccl28* primers were as follows: sense, acctcagaagccatactctcc; and antisense, tactcttgaggtctctcatcactgc.

**Western blotting, ELISAs and protein array analyses.** Ten frozen, stage III ovarian cancer samples, different from those used to generate the seven primary cell lines used in hypoxia experiments, were homogenized in lysis buffer (Pierce). Routine spectroscopic protein methods were used to determine the protein concentration, and 100 µg protein was loaded onto 8% SDS-PAGE gels, with the separated proteins subsequently transferred to Hybond membranes (Amersham Biosciences). The membranes were blocked with 10% skimmed milk and incubated with mouse anti-human CCL28 antibody (2.5 µg ml<sup>-1</sup>; R&D Systems, clone 62705) for 1 h, washed and then incubated with a goat anti-mouse horseradish-peroxidase-conjugated antibody (BD Pharmingen) for 45 min. Immunoreactive bands were detected using the ECL detection system (Amersham Pharmacia).

The DuoSet ELISA Development Kit (R&D Systems) was used, according to the manufacturer's instructions, to detect the following: human CCL28 in tumour cell supernatants; mouse CCL28 and VEGFA in mouse ascites and tumour cell culture supernatants; and mouse VEGFA, basic fibroblast growth factor (bFGF), hepatocyte growth factor (HGF) and placental growth factor (PIGF) in supernatants from hypoxic or oxic mouse T<sub>reg</sub> cells. Quantification of growth factors and cytokines in ascitic fluid was performed when animals reached a weight of ~30 g in each group. Ascites were collected from nine mice per group by using paracentesis. Ascites from sets of three mice were pooled to obtain three samples per group. Protein arrays were performed by the company Rules-Based Medicine using rodent Multi-Analyte Profiles (MAPs).

**Immunostaining.** Human CCL28, cytokeratins and HIF1α were detected by immunohistochemistry (IHC) using the human ovarian cancer tissue microarray described above. Sections were cut at a 5 µm thickness and singly immunostained with anti-cytokeratin (Dako, DAB chromogen), monoclonal mouse anti-human HIF1α (NeoMarkers MS-1164-P, DAB chromogen), monoclonal mouse anti-human CCL28 (R&D Systems MAB7171, Fast Red chromogen) or monoclonal

mouse anti-human FOXP3 (BioLegend 320102, DAB chromogen) antibody. Slides were scanned on a whole-slide imaging system (Aperio), and immunoreactivity was scored by a trained pathologist (I.S.H.). For HIF1 $\alpha$  and CCL28, a semiquantitative scale ranging from 0 (no reactivity) to 3 (strong reactivity) was used. A direct count of intraepithelial and total FOXP3<sup>+</sup> lymphocyte nuclei was also made. Cores with inadequate histology or without tumour were disregarded. For HIF1 $\alpha$ , 207 cores were countable of 264 on the array. For CCL28, 229 of 264 cores were interpretable. For FOXP3 (intraepithelial), 189 of 264 cores were interpretable, and for FOXP3 (total), 199 of 264 cores were interpretable.

Mouse CCL28, carbonic anhydrase IX (CA IX) and CD31 were detected by IHC in intraperitoneal ID8 and ID8-ccl28 tumour nodules of ~5 mm in average diameter that were removed from mice when mice in each group reached a weight of ~30 g. Tumours were embedded in OCT medium and immediately snap frozen in dry ice. Sections (6  $\mu$ m thickness) were stained for mouse CCL28 (R&D Systems MAB533, Perma Red chromogen), CA IX (R&D Systems AF2344, DAB chromogen) or CD31 (BD Pharmingen clone 390, 558737, DAB chromogen) using IHC. For quantitative analysis of the CCL28 and CA IX correlation, frozen sections of tumours were double immunostained with a polyclonal goat anti-mouse CA IX antibody and a monoclonal rat anti-mouse CCL28 antibody, with haematoxylin as a counterstain. A total of twelve  $\times$ 200 high-power fields were imaged for each of three tumour samples per mouse from three mice in each group, and the spectral components were deconvoluted using the Nuance FX multispectral imaging system (Caliper Life Sciences). Sixty-two regions of interest (ROIs) were designated manually around histologically intact areas containing approximately 20 nuclei each (Supplementary Fig. 13). These were arbitrarily designated without regard to CA IX or CCL28 staining. The mean DAB (CA IX) and Perma Red (CCL28) intensities were quantified in these ROIs using the Nuance FX multispectral imaging system. To assess the relationship between CA IX and CCL28, linear regression was performed with the program Prism 5 (GraphPad Software).

**Flow cytometry.** Cells were subjected to up to six-colour flow cytometry on a FACSCanto flow cytometer using CellQuest Pro 3.2.1f1 software (BD Biosciences); data were analysed using FlowJo (Tree Star). The following monoclonal antibodies against mouse markers were used: phycoerythrin (PE)-Cy7-conjugated anti-CD45, allophycocyanin (APC)-Cy7-conjugated anti-CD3, PE-conjugated anti-CD4, peridinin-chlorophyll-protein (PerCP)-conjugated anti-CD8, APC-conjugated anti-CCR10 (R&D Systems clone 248918), PerCP-conjugated anti-CCR3 (R&D Systems clone 83101), fluorescein isothiocyanate (FITC)-conjugated anti-CD25 (R&D Systems), the APC-conjugated anti-mouse/rat FOXP3 Staining Set (eBioscience clone FJK-16s), PerCP-conjugated anti-Gr1, PE-Cy7-conjugated anti-CD11b, APC-conjugated anti-CD14, FITC-conjugated anti-PDCA-1 and PE-conjugated anti-CD123 antibody. Where the clone number is not indicated, different clones were used with the same results. The following monoclonal antibodies against human markers were used: PerCP-conjugated anti-CD45, APC-Cy7-conjugated anti-CD4, PE-Cy7-conjugated anti-CD3 and FITC-conjugated anti-CD25 (BD Pharmingen) antibody, and the APC-conjugated anti-human FOXP3 Staining Set (eBioscience, clone PCH101). Where the clone number is not indicated, different clones were used with the same results. Experiments in animals were performed in five to eight animals per group and were repeated at least twice. Staining on cells was performed in triplicate in at least three independent experiments.

**Migration assays, T<sub>reg</sub>-cell staining, proliferation and MLRs.** Supernatants (150  $\mu$ l) from hypoxic or oxic human ovarian cancer cells (OVCAR5, PEO4 or SKOV3), or PBS (containing 1% FBS) with or without 1–2  $\mu$ g ml<sup>-1</sup> of human recombinant CCL28, were plated into the bottom of 5- $\mu$ m-pore migration chambers (Corning). In some experiments, medium or PBS solution was preincubated for 1 h with anti-CCL28 antibody (R&D Systems clone MAB717). One million fresh human PBMCs were seeded in 50  $\mu$ l PBS containing 1% FBS in the top of the migration chambers. In some experiments, PBMCs were previously incubated for 1 h with anti-CCR10 antibody (Abcam Ab12548), anti-CCR3 antibody (Abcam Ab25789) or IgG isotype control (R&D Systems 43414). Following ~4 h incubation at 37 °C, cells migrating to the lower chambers were collected and used for T<sub>reg</sub>-cell analysis, using antibodies against human CD45, CD3, CD4, CD25 and FOXP3 (BD Pharmingen), or for mixed leukocyte reactions (MLRs). Similar experiments were conducted with carboxy-fluorescein diacetate succinimidyl ester (CFSE)-labelled spleen T cells from healthy C57BL/6 mice that were seeded against fresh ascites from mice with ID8 or ID8-ccl28 tumours. For T<sub>reg</sub>-cell functional assays, 2  $\times$  10<sup>5</sup> CFSE-labelled spleen T cells that had migrated to the lower chambers containing ascites were collected and incubated for 5 days with dynabeads decorated with anti-CD3/anti-CD28 antibodies at one-tenth the optimal concentration

recommended by the manufacturer. In another experiment, cells derived from ID8 or ID8-ccl28 ascites were allowed to adhere to plastic for 3 h at 37 °C. Floating cells were collected and processed to purify tumour-associated T cells using a PAN T Cell Isolation Kit (Miltenyi). Target BALB/c splenocytes were irradiated at 3,000 cGy and were cultured in a 1:1 or 1:10 ratio with responder T cells from ID8 or ID8-ccl28 ascites in RPMI-1640 containing 10% heat-inactivated FCS, 2 mM glutamine, 1 mM sodium pyruvate, 100 U ml<sup>-1</sup> penicillin, 100 U ml<sup>-1</sup> streptomycin and 5  $\mu$ g ml<sup>-1</sup> gentamicin sulphate. In proliferation assays, cells were titrated, with normalization based on the frequency of responder cells such that 1  $\times$  10<sup>5</sup> CD3<sup>+</sup>FOXP3<sup>-</sup> (responder) T cells per well were cultured with 1  $\times$  10<sup>4</sup> to 1  $\times$  10<sup>5</sup> irradiated BALB/c splenocytes per well in 96-well, round-bottom plates. Incorporation of [<sup>3</sup>H]thymidine was assessed during the last 16 h of culture. In all experiments using mouse splenocytes or ascites-derived T cells, cells were freshly processed. Following collection, cells were centrifuged and rinsed twice to remove tissue or fluid debris and were then used for *in vitro* experiments.

**Endothelial-tube formation assay and *in vivo* angiogenesis.** Pools of human umbilical vein endothelial cells (HUVECs, Cambrex) were grown in reduced (VEGF-free) EBM-2/EGM-2 medium (Cambrex). Human CD4<sup>+</sup>CD25<sup>+</sup> T<sub>reg</sub> cells were purified using a magnetic activated cell sorting (MACS) T<sub>reg</sub>-cell purification kit (Miltenyi). Purified T<sub>reg</sub> cells and T<sub>reg</sub>-cell-depleted CD4<sup>+</sup> cells were incubated for 24 h under hypoxic or oxic conditions, as above. The tube formation assay was performed as previously described<sup>35,36</sup>. Briefly, growth-factor-reduced Matrigel (BD Biosciences, 250  $\mu$ l well<sup>-1</sup>) was allowed to polymerize in a 24-well plate at 37 °C for at least 30 min. HUVECs (5  $\times$  10<sup>4</sup> cells well<sup>-1</sup>) were suspended in 250  $\mu$ l medium conditioned by oxic or hypoxic T<sub>reg</sub> cells or T<sub>reg</sub>-cell-depleted CD4<sup>+</sup> cells in the presence or absence of neutralizing antibody specific for VEGF receptor 1 (VEGFR1, 10  $\mu$ g ml<sup>-1</sup>, R&D Systems clone 49560) and VEGFR2 (10  $\mu$ g ml<sup>-1</sup>, R&D Systems clone 89106). After incubation for 24 h at 37 °C, capillary-like structures in Matrigel were photographed under a phase contrast microscope. Total tube length was quantified using the image analysis software Image-Pro Plus (version 3.0, Media Cybernetics).

*In vivo* angiogenesis assays were performed by mixing 100  $\mu$ l medium conditioned by CD4<sup>+</sup>CD25<sup>+</sup> or CD4<sup>+</sup>CD25<sup>-</sup> cells with 200  $\mu$ l growth-factor-free Matrigel. The mixture was injected subcutaneously in mice. Matrigel plugs were removed 72 h after implantation, dispersed by using a cell strainer and centrifuged at 1,200 r.p.m. for 10 min. The collected cells were resuspended and then analysed by flow cytometry for cell-surface CD31 and CD45, as above.

**Biocomputational and statistical methods.** *P* values associated with all pairwise comparisons were based on Student's *t*-test for independent groups. No adjustments for multiple hypothesis tests were made. Error bars were defined using standard deviation for all *in vitro* experiments and the standard error of the mean for all tumour measurements *in vivo*, except where noted in the figure legend. The median (HIF1 $\alpha$  and CCL28) or mean (FOXP3) scores in IHC were correlated using a non-parametric method (Spearman's rho, two-tailed) in the program Prism 5. To assess the relationship between CA IX and CCL28, linear regression was performed in Prism 5.

Two publicly available Affymetrix array expression data sets (GSE3149 and GSE9891), comprising samples from a total of 353 human ovarian cancer patients from Duke University<sup>31</sup> and the Australian Ovarian Cancer Study<sup>32</sup>, respectively, were analysed. CEL files were downloaded from the Gene Expression Omnibus database (GEO, National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE9899>). Data were processed using the Robust Multichip Average (RMA)<sup>37</sup>, and expression values less than zero were given a value of 0.01. An optimal cut-off point defining two groups of patients with different survival curves using CCL28 gene expression (CCL28\_1 probe) was determined using the program X-tile<sup>33</sup>. Kaplan–Meier curves were computed, and the log-rank test was used to determine whether the survival curves were significantly different.

34. Bergeron, M. *et al.* Detection of hypoxic cells with the 2-nitroimidazole, EF5, correlates with early redox changes in rat brain after perinatal hypoxia–ischemia. *Neuroscience* **89**, 1357–1366 (1999).
35. Lee, O. H. *et al.* Sphingosine 1-phosphate induces angiogenesis: its angiogenic action and signaling mechanism in human umbilical vein endothelial cells. *Biochem. Biophys. Res. Commun.* **264**, 743–750 (1999).
36. Lee, Y. H. *et al.* Cell-retained isoforms of vascular endothelial growth factor (VEGF) are correlated with poor prognosis in osteosarcoma. *Eur. J. Cancer* **35**, 1089–1093 (1999).
37. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).



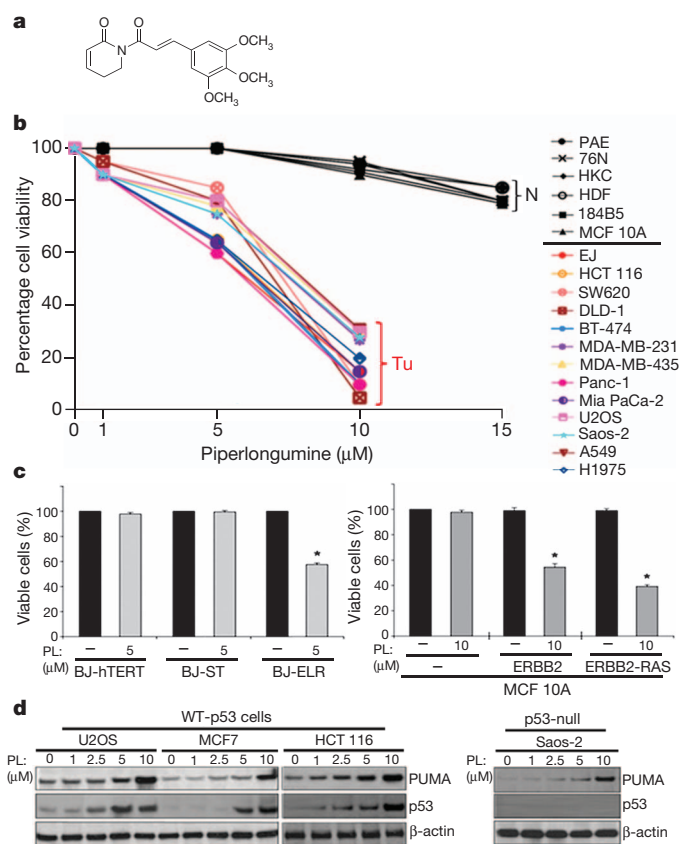
# Selective killing of cancer cells by a small molecule targeting the stress response to ROS

Lakshmi Raj<sup>1</sup>, Takao Ide<sup>1</sup>, Aditi U. Gurkar<sup>1</sup>, Michael Foley<sup>2</sup>, Monica Schenone<sup>2</sup>, Xiaoyu Li<sup>2</sup>, Nicola J. Tolliday<sup>2</sup>, Todd R. Golub<sup>2</sup>, Steven A. Carr<sup>2</sup>, Alykhan F. Shamji<sup>2</sup>, Andrew M. Stern<sup>2</sup>, Anna Mandinova<sup>1,2</sup>, Stuart L. Schreiber<sup>2</sup> & Sam W. Lee<sup>1,2</sup>

**Malignant transformation, driven by gain-of-function mutations in oncogenes and loss-of-function mutations in tumour suppressor genes, results in cell deregulation that is frequently associated with enhanced cellular stress (for example, oxidative, replicative, metabolic and proteotoxic stress, and DNA damage)<sup>1</sup>. Adaptation to this stress phenotype is required for cancer cells to survive, and consequently cancer cells may become dependent upon non-oncogenes that do not ordinarily perform such a vital function in normal cells. Thus, targeting these non-oncogene dependencies in the context of a transformed genotype may result in a synthetic lethal interaction and the selective death of cancer cells<sup>2</sup>. Here we used a cell-based small-molecule screening and quantitative proteomics approach that resulted in the unbiased identification of a small molecule that selectively kills cancer cells but not normal cells. Piperlongumine increases the level of reactive oxygen species (ROS) and apoptotic cell death in both cancer cells and normal cells engineered to have a cancer genotype, irrespective of p53 status, but it has little effect on either rapidly or slowly dividing primary normal cells. Significant antitumour effects are observed in piperlongumine-treated mouse xenograft tumour models, with no apparent toxicity in normal mice. Moreover, piperlongumine potently inhibits the growth of spontaneously formed malignant breast tumours and their associated metastases in mice. Our results demonstrate the ability of a small molecule to induce apoptosis selectively in cells that have a cancer genotype, by targeting a non-oncogene co-dependency acquired through the expression of the cancer genotype in response to transformation-induced oxidative stress<sup>3–5</sup>.**

Using a luciferase reporter gene fused with the *CDIP* (cell death involved p53 target, also known as 5730403B10Rik) promoter<sup>6</sup>, we performed a small-molecule screen (Supplementary Fig. 1) to identify compounds acting through novel pro-apoptotic mechanisms. The compound with the highest composite Z value was piperlongumine (Supplementary Fig. 2a), which increased luciferase activity from the reporter gene at levels comparable to the positive control, etoposide (Supplementary Figs 2b and 3). Piperlongumine is a natural product isolated from the plant species *Piper longum* L. (Fig. 1a) and it was previously shown to have cytotoxic effects<sup>7</sup>. We examined the effects of piperlongumine on the viability of cultured cancer cells and normal cells (Fig. 1b and Supplementary Figs 4 and 6). Piperlongumine treatment markedly induced cell death in cancer cells with both wild-type p53 and mutant p53. When primary normal cells and non-transformed immortalized cells with diverse proliferative capacities were incubated with highly purified piperlongumine (Supplementary Fig. 5) for 24 h (under the indicated conditions, which avoid spontaneous transformation and minimize stress), there was little apparent reduction in cell viability, even at the highest concentration tested (15  $\mu$ M, a concentration of piperlongumine that approaches its solubility limit). This indicated that piperlongumine may have a cancer-cell-selective killing property, and that sensitivity to piperlongumine may result from the process of malignant transformation. To test this hypothesis, we used a

defined model<sup>8</sup> of oncogenic conversion of normal cells through ectopic expression of the telomerase catalytic subunit (*hTERT*) in combination with small T antigen and an oncogenic allele of *HRAS* (Fig. 1c), and



**Figure 1 | Selective killing effect of piperlongumine in cancer cells.**

**a**, Structure of piperlongumine. **b**, Piperlongumine treatment induces cell death in cancer cells but not in normal cells. Normal human cells (N), including aortic endothelial cells (PAE), breast epithelial cells (76N), keratinocytes (HKC) and skin fibroblasts (HDF), as well as two immortalized breast epithelial cell lines (184B5 and MCF 10A), were grown in 12-well or 24-well plates and treated with piperlongumine at 1–15  $\mu$ M for 24 h. A variety of human cancer cell lines (Tu) were also treated with piperlongumine or DMSO (control) for 24 h. Cytotoxicity was measured by trypan blue exclusion staining (average of three independent experiments). Piperlongumine was HPLC-purified (~99% purity) before the treatment. **c**, Selective cell death caused by piperlongumine (PL) in oncogenically transformed human BJ skin fibroblasts (left panel) and MCF 10A cell lines (right panel). A representative graph for cell viability is shown (mean  $\pm$  s.d. of three independent experiments; \*,  $P < 0.0001$ ). **d**, The effects of piperlongumine on p53 and its target PUMA were measured by western blot analyses in several cancer cell lines.  $\beta$ -actin expression was used as a loading control.

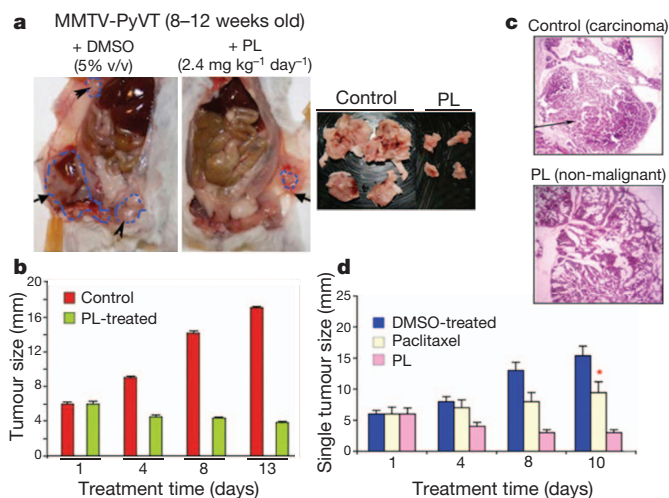
<sup>1</sup>Cutaneous Biology Research Center, Massachusetts General Hospital and Harvard Medical School, Building 149 13<sup>th</sup> Street, Charlestown, Massachusetts 02129, USA. <sup>2</sup>Broad Institute of Harvard and MIT, 7 Cambridge Center, Massachusetts 02142, USA.

observed sensitivity to piperlongumine upon oncogenic transformation of normal cells. Similar results were obtained using serial transformation of spontaneously immortalized MCF 10A breast epithelial cells by overexpression of *ERBB2* and/or *HRAS*<sup>9</sup> (Fig. 1c).

Western blot analysis showed that wild-type p53 expression was significantly enhanced in different types of cancer cells by treatment with piperlongumine (Fig. 1d). Moreover, a p53 proapoptotic target, BCL2 binding component 3 (BBC3, also known as PUMA), was significantly induced in response to piperlongumine, even in p53-null Saos-2 cancer cells (Fig. 1d). Piperlongumine treatment was able to repress the expression of several pro-survival proteins, including B-cell CLL/lymphoma 2 (BCL2), baculoviral IAP repeat containing 5 (also known as survivin) and X-linked inhibitor of apoptosis (XIAP) (Supplementary Fig. 7). Among 55 death- or survival-related genes, we observed increased levels of apoptotic transcripts and decreased levels of pro-survival transcripts in cancer cells in the presence of piperlongumine, but no significant changes in these transcripts in normal cells (Supplementary Fig. 8). These results indicate that piperlongumine induces cell death or apoptosis (Supplementary Fig. 4a, e) preferentially in cancer cells by modulating the expression of members of apoptotic and survival pathways, including p53 targets and p53 itself, and that it does not require p53 for this activity.

We next tested piperlongumine in established tumour xenografts in mice (human bladder, breast and lung tumours in nude mice, and mouse melanoma in C57BL/6 mice; Supplementary Fig. 9). Marked antitumour effects were observed in tumour-bearing mice treated with piperlongumine, as compared to dimethyl sulphoxide (DMSO)-treated controls (Supplementary Fig. 9). Piperlongumine treatment enhanced the expression of cyclin-dependent kinase inhibitor 1A (CDKN1A, or p21<sup>WAF1/CIP1</sup>), PUMA and caspase 3 in EJ-cell tumours (Supplementary Fig. 10a). Moreover, piperlongumine treatment inhibited the formation of blood vessels in xenograft-tumour mice (Supplementary Figs 9d and 10b). We also studied piperlongumine in a transgenic mouse model of spontaneous breast cancer, MMTV-PyVT<sup>10</sup>. When tumour sizes had grown to about 5–6 mm in diameter (in female MMTV-PyVT mice, 8–9 weeks of age), piperlongumine was administered intraperitoneally (2.4 mg kg<sup>-1</sup>) daily for two weeks and notable antitumour effects were observed (Fig. 2a, b). Furthermore, there were no secondary tumours in piperlongumine-treated mice compared to vehicle-treated controls. At day 13, the vehicle-treated control mice showed severe malignant progression indicated by the formation of aggressive adenocarcinoma (Fig. 2c). In contrast, the mammary glands of piperlongumine-treated mice were preserved and the tissue showed a hyperplastic-like, non-malignant phenotype (Fig. 2c). Notably, piperlongumine seemed to be more effective in tumour growth inhibition than paclitaxel (Fig. 2d). Piperlongumine also showed excellent oral bioavailability and desirable exposure levels ( $C_{\max}$  and bioavailability, measured by calculating the area under curve (AUC)) in mice, as observed after a single oral administration and after intravenous injection (Supplementary Fig. 11). To examine potential cytotoxic side-effects of piperlongumine on normal tissues, CD-1 mice were intraperitoneally treated with piperlongumine (2.4 mg kg<sup>-1</sup>) or DMSO, daily for 6 days, and whole blood samples as well as vital organs were collected for haematology and histopathological analyses, respectively. Piperlongumine-treated CD-1 mice remained healthy throughout the treatment time and no notable differences between the vehicle-treated and piperlongumine-treated groups were evident (Supplementary Figs 12 and 13). High-dose acute toxicity studies demonstrated that piperlongumine did not cause any obvious clinical indications (Supplementary Fig. 14 and Supplementary Tables 1 and 2). Together, these results indicate that treatment with piperlongumine potently suppresses tumour growth in diverse tissues without affecting normal tissues in mice.

We next used a method combining affinity enrichment with stable-isotope labelling with amino acids in cell culture (SILAC) and quantitative proteomics to identify the target proteins and their associated

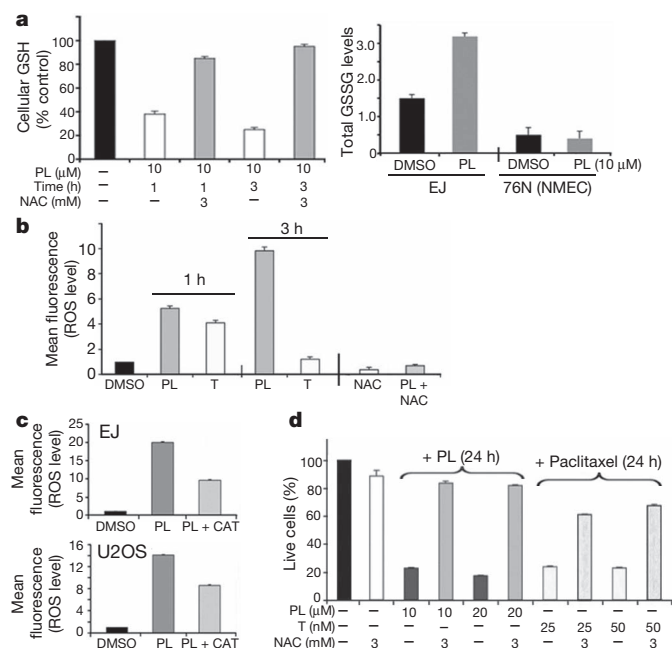


**Figure 2 | In vivo antitumour effect of piperlongumine.** **a**, Inhibition of mammary tumour growth by piperlongumine treatment in MMTV-PyVT transgenic tumour-bearing mice. MMTV-PyVT mice spontaneously developed breast adenocarcinoma by ~8 weeks of age. When tumours had grown to ~5–6 mm in diameter, piperlongumine (2.4 mg kg<sup>-1</sup>), paclitaxel (10 mg kg<sup>-1</sup>) or DMSO (5% v/v) was administered intraperitoneally daily for 13 days ( $n = 12$  mice per group). Mice were then euthanized and mammary tumours were excised and processed for histological examination. Arrows indicate the initial as well as secondary tumour lesions in the DMSO treated animals (left panel) versus only a single initial lesion in the piperlongumine treated animals (right panel). **b**, The sizes of the grossly dissected tumours were measured and plotted. **c**, Histological morphology of mammary tissue sections from MMTV-PyVT tumour-bearing mice treated with piperlongumine or DMSO after 13 days, stained with haematoxylin and eosin. **d**, Size of single tumours after 10 days of piperlongumine or paclitaxel treatment (asterisk indicates shorter treatment period due to high toxicity of paclitaxel in animals). Values in bar graphs are mean  $\pm$  s.d. of three independent experiments.

complexes that bind to piperlongumine<sup>11</sup> (Supplementary Figs 15 and 16a and Supplementary Methods). Twelve interaction partners of piperlongumine that were similar in both EJ and U2OS cells were identified (Supplementary Fig. 16b). Seven of these, including the top four high-signal outliers, are known to participate in the cellular response to oxidative stress caused by elevated ROS. Glutathione S-transferase pi 1 (GSTP1) was the highest-confidence hit, followed by carbonyl reductase 1 (CBR1) (Supplementary Fig. 16b). Several of these proteins are known to be part of a common complex<sup>12,13</sup>, indicating that the affinity purification may have identified direct and indirect partners.

These results indicate that, by binding to proteins known to regulate oxidative stress, piperlongumine may modulate redox and ROS homeostasis. Consistent with this hypothesis, we found that piperlongumine can interact directly with purified recombinant GSTP1 and inhibit its activity (Supplementary Figs 17 and 18), and also that it can lead to a decrease in reduced glutathione (GSH) levels and an increase in oxidized glutathione (GSSG) levels in cancer cells (Fig. 3a). Piperlongumine treatment did not increase GSSG levels in normal cells (76N (NMEC)) (Fig. 3a). Furthermore, co-treatment with piperlongumine and the reducing agent *N*-acetyl-L-cysteine (NAC, 3 mM), which quenches ROS, prevented piperlongumine-mediated GSH depletion (Fig. 3a).

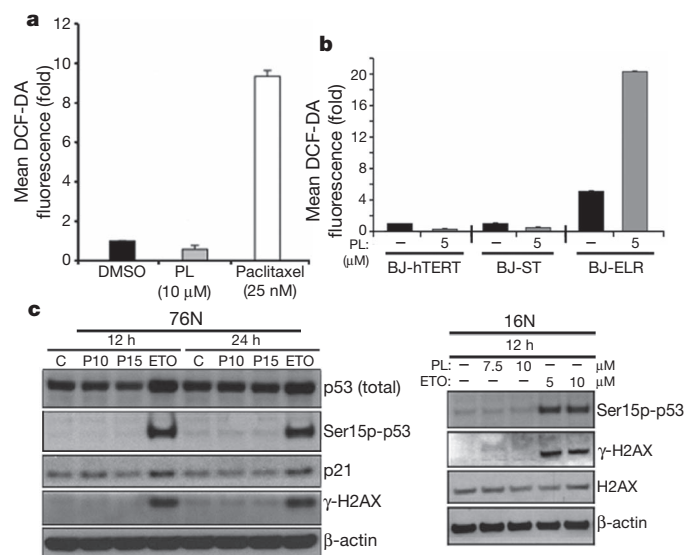
We next determined the effect of piperlongumine on cellular ROS levels in several human cancer cells (EJ, MDA-MB-231, U2OS and MDA-MB-435) through flow cytometry using the redox-sensitive fluorescent probe 2',7'-dichlorofluorescein diacetate (DCF-DA). Treatment with piperlongumine for 1 h and 3 h caused a marked increase in ROS levels in these cancer cells (Fig. 3b and Supplementary Figs 19 and 20). Paclitaxel also caused an increase in DCF-DA fluorescence after 1 h, but piperlongumine enhanced ROS to nearly



**Figure 3 | Piperlongumine enhances ROS accumulation in cancer cells by targeting the stress response to ROS.** **a**, Piperlongumine-mediated modulation of GSH and GSSG. GSH levels were determined after EJ cells were either treated with piperlongumine or pretreated with NAC for 1 h, followed by piperlongumine treatment for 1 h or 3 h (left panel). GSSG levels were also determined after EJ cells and 76N (NMEC) cells were treated with piperlongumine for 3 h (right panel). **b**, Piperlongumine-induced ROS elevation and reversion by NAC. EJ cells were treated with piperlongumine (PL, 10  $\mu$ M), paclitaxel (T, 25 nM) or DMSO for 1 h and 3 h. Cells were also pretreated with 3 mM NAC for 1 h, followed by 10  $\mu$ M piperlongumine for 3 h. **c**, Reversion of piperlongumine-induced ROS accumulation by catalase. EJ or U2OS cells were pretreated with catalase (CAT, 2,000 U ml<sup>-1</sup>) for 2 h, followed by 10  $\mu$ M piperlongumine for 3 h. **d**, Piperlongumine-induced cell death can be rescued by NAC. EJ cells were treated with piperlongumine for 24 h, or treated with 3 mM NAC for 1 h followed by piperlongumine or paclitaxel for 24 h. Cell viability was measured by trypan blue exclusion staining assay. All values are mean  $\pm$  s.d. of three independent experiments.

twice these levels (Fig. 3b). Co-treatment with NAC fully reversed the piperlongumine-induced increase in ROS and cell death (Fig. 3b, d and Supplementary Fig. 21). Using a series of fluorescent probes specific for individual species of ROS, we found that hydrogen peroxide and nitric oxide, but not superoxide anion, were among the ROS species induced by piperlongumine in cancer cells (Fig. 3b, c and Supplementary Figs 22–24).

In contrast to the results in cancer cells, piperlongumine did not cause an increase in ROS levels in normal cells (Fig. 4a and Supplementary Fig. 25). This selective induction of ROS in cancer cells distinguishes piperlongumine from other small molecules that affect ROS levels, such as the microtubule-stabilizing agent paclitaxel and the glutathione synthesis inhibitor buthionine sulfoximine (Fig. 4a and Supplementary Fig. 25), and indicates that piperlongumine-induced ROS elevation is a consequence of cell transformation. Engineering normal cells to have a cancer genotype potentiated the piperlongumine-induced increase in ROS (Fig. 4b and Supplementary Figs 26 and 27). Serial transformation itself leads to increased expression of the putative piperlongumine targets GSTP1 and CBR1 (Supplementary Fig. 28), indicating that these proteins may have a role in enabling the transformed cell to adapt to transformation-induced oxidative stress. We therefore hypothesized that overexpression of CBR1 or GSTP1 might rescue transformed cells from both piperlongumine-induced ROS elevation and piperlongumine-induced apoptosis. Stably overexpressing CBR1 or GSTP1, and particularly both, in EJ cells markedly reduced piperlongumine-induced ROS levels and partially rescued



**Figure 4 | Piperlongumine does not increase ROS levels and the ROS-induced DNA-damage response in normal and immortalized non-transformed cells.** **a**, Piperlongumine does not increase ROS levels in normal cells (16N). ROS levels were measured by flow cytometry and shown by quantitative bar graph measured as the fold change over DMSO-treated levels. **b**, Selective induction of ROS by piperlongumine in oncogenically transformed BJ human fibroblasts (BJ-ELR), but not in non-transformed BJ fibroblasts (BJ-hTERT and BJ-ST). ROS levels were measured after treating with DMSO (–) or piperlongumine (PL, 5  $\mu$ M) for 8 h. All values are mean  $\pm$  s.d. of three independent experiments. **c**, The effect of piperlongumine on stress-response targets was determined by western blot analysis of p53, phosphorylated p53 (Ser 15 p-p53), p21 and  $\gamma$ -H2AX in normal 76N cells, treated with piperlongumine (10  $\mu$ M (P10) or 15  $\mu$ M (P15)), 10  $\mu$ M etoposide (ETO) or DMSO as a solvent control (C) for 12 h and 24 h (left panel). Western blots were performed similarly on 16N cells treated for 12 h with piperlongumine (7.5  $\mu$ M or 10  $\mu$ M) or etoposide (5  $\mu$ M or 10  $\mu$ M) (right panel).  $\beta$ -actin expression was used as a loading control.

the piperlongumine-induced apoptotic phenotype (Supplementary Fig. 29). In a complementary study, knockdown of GSTP1 or CBR1 did not affect piperlongumine-induced ROS levels (Supplementary Fig. 30). These results may reflect the fact that other members of the GST family were observed to bind piperlongumine in our affinity-enrichment studies (Supplementary Fig. 16b) and may have partially overlapping functions in the cell. These data indicate that piperlongumine induces apoptosis by interfering with redox and ROS homeostatic regulators such as GSTP1 and CBR1.

The ability of piperlongumine to inhibit the growth of rapidly growing and highly invasive multifocal mammary tumours without general toxicity indicates that perturbing redox and ROS homeostasis is a promising strategy for cancer treatment. Our cell-based experiments indicate that piperlongumine treatment selectively increases ROS levels and induces apoptosis in cancer cells relative to normal cells. This correlates with the selective induction of related phenotypes, including DNA damage (Fig. 4c and Supplementary Figs 27, 31 and 32) and alterations in mitochondrial morphology and function, occurring selectively in cancer cells (Supplementary Fig. 33). The differential response of cancer cells and normal cells to treatment with piperlongumine indicates that piperlongumine targets a dependency associated with ROS homeostasis that arises during transformation. Normal cells, including stem cells, have low basal levels of ROS<sup>1,4,5,14–17</sup> and therefore a diminished reliance on the ROS stress-response pathway, whereas cancer cells, especially cancer stem cells, have high levels of ROS<sup>14</sup> and might therefore be expected to have a strong reliance on the ROS stress-response pathway<sup>1,5,16,18,19</sup>. The use of small molecules that alter levels of ROS such as  $\beta$ -phenylethyl-isothiocyanate and buthionine sulfoximine<sup>4,17</sup> has been suggested for the treatment of cancer. Other small molecules such as curcumin<sup>20</sup> and 2-cyano-3, 12-dioxoolean-1,9-dien-28-oic acid (CDDO) derivatives<sup>21</sup> have been



reported to promote ROS and reduce GSH levels in cancer cells, in one case in an oncogene-dependent manner<sup>17</sup>, and the activation of the KEAP1–NRF2 antioxidant pathway<sup>22,23</sup> has been suggested to be involved.

The introduction of a single oncogene (*HRAS*) leads to increased levels of ROS (Fig. 4b and ref. 24), increased expression of GSTP1 and CBR1 (Supplementary Fig. 28), an increased apoptotic response to piperlongumine (Fig. 1c), and notably, to a substantial increase in levels of ROS after treatment with piperlongumine. In EJ cells, piperlongumine-induced cell death is rescued by the antioxidant NAC (Fig. 3d). The increased dependence of cancer cells on the ROS stress-response pathway may be the basis for the selectivity of piperlongumine-induced apoptosis in cancer cells (Figs 1 and 2). In support of this hypothesis, the activation of signalling through the JNK (also known as MAPK8) pathway has been implicated as an antitumorigenic response to oncogene expression<sup>25</sup>. This response is coupled to oncogene-dependent oxidative stress through p53 stabilization, and could also function independently of p53 through pro-apoptotic cJUN-dependent transcription. In addition to its role in regulating ROS, GSTP1 is also known to be a direct negative regulator of JNK<sup>26</sup>, providing a possible mechanism for piperlongumine-induced apoptosis in both p53-wild-type and p53-mutant cancer cells.

A global investigation of the spectrum of cancer genotypes will be required to identify the range of cancer genotypes that impart piperlongumine sensitivity, but our results already highlight a novel strategy for cancer therapy that preferentially eradicates cancer cells by targeting the ROS stress-response pathway.

## METHODS SUMMARY

Apoptotic cell populations were determined by TdT-mediated dUTP nick end labelling (TUNEL) assay and quantified using flow cytometry. Cell viability was also determined by crystal violet staining (0.2% w/v in 2% ethanol), by trypan blue exclusion and by the Alamar blue cell viability assay. For crystal violet staining, cells were plated in 6-well and 12-well plates and, after reaching 60–70% confluency, the cells were treated with piperlongumine for 12 h and 24 h. For measurement of ROS production, cells were treated with piperlongumine or paclitaxel for 1 h and 3 h and then incubated with 10  $\mu$ M DCF-DA for 30 min at 37 °C, washed twice with PBS and immediately analysed by a FACScan flow cytometer. Cells were treated with piperlongumine and etoposide for 18–24 h and processed for Comet assay following the manufacturer's instructions (Trevigen). For xenograft tumour models, cancer cell lines EJ, A549 and MDA-MB-435 were injected subcutaneously into the flanks of nude mice. For the melanoma mouse model, B16-F10 melanoma cells were injected into the flanks of C57BL/6J mice. FVB/N-Tg (MMTV-PyVT) 634Mul males were obtained from the Mouse Models of Human Cancer consortium (MMHCC) at NCI-Frederick and bred with FVB females. Female offspring were genotyped for the presence of the transgene using the primers published by MMHCC. For piperlongumine target identification, we followed the SILAC-based affinity enrichment methodology previously described<sup>11,27</sup>. For further details see Supplementary Methods.

Received 17 February 2010; accepted 3 May 2011.

1. Luo, J., Solimini, N. L. & Elledge, S. J. Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell* **136**, 823–837 (2009).
2. Yap, T. A., Sandhu, S. K., Carden, C. P. & de Bono, J. S. Poly(ADP-ribose) polymerase (PARP) inhibitors: Exploiting a synthetic lethal strategy in the clinic. *CA Cancer J. Clin.* **61**, 31–49 (2011).
3. Poole, L. B. & Nelson, K. J. Discovering mechanisms of signaling-mediated cysteine oxidation. *Curr. Opin. Chem. Biol.* **12**, 18–24 (2008).
4. Schumacker, P. T. Reactive oxygen species in cancer cells: live by the sword, die by the sword. *Cancer Cell* **10**, 175–176 (2006).
5. Trachootham, D., Alexandre, J. & Huang, P. Targeting cancer cells by ROS-mediated mechanisms: a radical therapeutic approach? *Nat. Rev.* **8**, 579–591 (2009).
6. Brown, L. et al. CDIP, a novel pro-apoptotic gene, regulates TNF $\alpha$ -mediated apoptosis in a p53-dependent manner. *EMBO J.* **26**, 3410–3422 (2007).

7. Bezerra, D. P. et al. Piplartine induces inhibition of leukemia cell proliferation triggering both apoptosis and necrosis pathways. *Toxicol. In Vitro* **21**, 1–8 (2007).
8. Hahn, W. C. et al. Creation of human tumour cells with defined genetic elements. *Nature* **400**, 464–468 (1999).
9. Ryo, A. et al. *PIN1* is an E2F target gene essential for *Neu/Ras*-induced transformation of mammary epithelial cells. *Mol. Cell. Biol.* **22**, 5281–5295 (2002).
10. Guy, C. T., Cardiff, R. D. & Muller, W. J. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol. Cell. Biol.* **12**, 954–961 (1992).
11. Ong, S. E. et al. Identifying the proteins to which small-molecule probes and drugs bind in cells. *Proc. Natl Acad. Sci. USA* **106**, 4617–4622 (2009).
12. Bateman, R. L., Rauh, D., Tavshanjian, B. & Shokat, K. M. Human carbonyl reductase 1 is an S-nitrosogluthione reductase. *J. Biol. Chem.* **283**, 35756–35762 (2008).
13. Ralat, L. A., Manevich, Y., Fisher, A. B. & Colman, R. F. Direct evidence for the formation of a complex between 1-cysteine peroxiredoxin and glutathione S-transferase  $\pi$  with activity changes in both enzymes. *Biochemistry* **45**, 360–372 (2006).
14. Diehn, M. et al. Association of reactive oxygen species levels and radioresistance in cancer stem cells. *Nature* **458**, 780–783 (2009).
15. Fruehauf, J. P. & Meyskens, F. L. Jr. Reactive oxygen species: a breath of life or death? *Clin. Cancer Res.* **13**, 789–794 (2007).
16. Huang, P., Feng, L., Oldham, E. A., Keating, M. J. & Plunkett, W. Superoxide dismutase as a target for the selective killing of cancer cells. *Nature* **407**, 390–395 (2000).
17. Trachootham, D. et al. Selective killing of oncogenically transformed cells through a ROS-mediated mechanism by  $\beta$ -phenylethyl isothiocyanate. *Cancer Cell* **10**, 241–252 (2006).
18. Gogvadze, V., Orrenius, S. & Zhivotovsky, B. Mitochondria in cancer cells: what is so special about them? *Trends Cell Biol.* **18**, 165–173 (2008).
19. Szatrowski, T. P. & Nathan, C. F. Production of large amounts of hydrogen peroxide by human tumor cells. *Cancer Res.* **51**, 794–798 (1991).
20. Ravindran, J., Prasad, S. & Aggarwal, B. B. Curcumin and cancer cells: how many ways can curry kill tumor cells selectively? *Am. Assoc. Pharm. Sci. J.* **11**, 495–510 (2009).
21. Yue, P., Zhou, Z., Khuri, F. R. & Sun, S. Y. Depletion of intracellular glutathione contributes to JNK-mediated death receptor 5 upregulation and apoptosis induction by the novel synthetic triterpenoid methyl-2-cyano-3, 12-dioxooleana-1, 9-dien-28-oate (CDDO-Me). *Cancer Biol. Ther.* **5**, 492–497 (2006).
22. Banning, A., Deubel, S., Kluth, D., Zhou, Z. & Brigelius-Flohe, R. The Gl-GPx gene is a target for Nrf2. *Mol. Cell. Biol.* **25**, 4914–4923 (2005).
23. Dinkova-Kostova, A. T. et al. An exceptionally potent inducer of cytoprotective enzymes: elucidation of the structural features that determine inducer potency and reactivity with KEAP1. *J. Biol. Chem.* **285**, 33747–33755 (2010).
24. Lee, A. C. et al. Ras proteins induce senescence by altering the intracellular levels of reactive oxygen species. *J. Biol. Chem.* **274**, 7936–7940 (1999).
25. Schramek, D. et al. The stress kinase MKK7 couples oncogenic stress to p53 stability and tumor suppression. *Nature Genet.* **43**, 212–219 (2011).
26. Wang, T., Arifoglu, P., Ronai, Z. & Tew, K. D. Glutathione S-transferase P1-1 (GSTP1-1) inhibits c-Jun N-terminal kinase (JNK1) signaling through interaction with the C terminus. *J. Biol. Chem.* **276**, 20999–21003 (2001).
27. Margolin, A. A. et al. Empirical Bayes analysis of quantitative proteomics experiments. *PLoS ONE* **4**, e7454 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank K. Chu, L. Brown-Endres, E. Lerner and F. Neville for their help in preparing the manuscript, W. C. Hahn for BJ cell lines, V. Band for 76N cells, D. Beer for H1975 cells and K. Todorova, G. Wei, S. Ong, S. Norton and F. An for technical assistance. This project has been supported in part by grants CA142805, CA127247, CA085681 and CA080058 from NIH. This research was supported by the National Cancer Institute's Initiative for Chemical Genetics Contract (N01-CO-12400) and Cancer Target Discovery and Development Network grant (5 RC2 CA148399-02), as well as the National Institutes of Health Genomics Based Drug Discovery—Target ID Project Grant (RL1HG004671, which is administratively linked to National Institutes of Health Grants RL1CA133834, RL1GM084437 and UL1RR024924). S.L.S. is an Investigator with the Howard Hughes Medical Institute.

**Author Contributions** L.R. and T.I. conducted most of the experimental work. A.U.G., M.S. and X.L. made critical experimental contributions. N.J.T., A.M.S., T.R.G., S.A.C., A.F.S. and M.F. designed the experimental plans, analysed and interpreted the data. A.M., S.L.S. and S.W.L. designed and directed the project and drafted the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.W.L. ([swlee@partners.org](mailto:swlee@partners.org)), A.M. ([amandinova@partners.org](mailto:amandinova@partners.org)) or S.L.S. ([stuart\\_schreiber@harvard.edu](mailto:stuart_schreiber@harvard.edu)).

# Low-energy control of electrical turbulence in the heart

Stefan Luther<sup>1,2,3,4\*</sup>, Flavio H. Fenton<sup>1,2\*</sup>, Bruce G. Kornreich<sup>5</sup>, Amgad Squires<sup>1,6</sup>, Philip Bittihn<sup>1,3</sup>, Daniel Hornung<sup>1,3</sup>, Markus Zabel<sup>4,7</sup>, James Flanders<sup>5</sup>, Andrea Gladuli<sup>5</sup>, Luis Campoy<sup>5</sup>, Elizabeth M. Cherry<sup>1,2,8</sup>, Gisa Luther<sup>1,3</sup>, Gerd Hasenfuss<sup>4,7</sup>, Valentin I. Krinsky<sup>1,9</sup>, Alain Pumir<sup>10</sup>, Robert F. Gilmour Jr<sup>2</sup> & Eberhard Bodenschatz<sup>1,3,4,6,11</sup>

**Controlling the complex spatio-temporal dynamics underlying life-threatening cardiac arrhythmias such as fibrillation is extremely difficult, because of the nonlinear interaction of excitation waves in a heterogeneous anatomical substrate<sup>1–4</sup>. In the absence of a better strategy, strong, globally resetting electrical shocks remain the only reliable treatment for cardiac fibrillation<sup>5–7</sup>. Here we establish the relationship between the response of the tissue to an electric field and the spatial distribution of heterogeneities in the scale-free coronary vascular structure. We show that in response to a pulsed electric field,  $E$ , these heterogeneities serve as nucleation sites for the generation of intramural electrical waves with a source density  $\rho(E)$  and a characteristic time,  $\tau$ , for tissue depolarization that obeys the power law  $\tau \propto E^\alpha$ . These intramural wave sources permit targeting of electrical turbulence near the cores of the vortices of electrical activity that drive complex fibrillatory dynamics. We show *in vitro* that simultaneous and direct access to multiple vortex cores results in rapid synchronization of cardiac tissue and therefore, efficient termination of fibrillation. Using this control strategy, we demonstrate low-energy termination of fibrillation *in vivo*. Our results give new insights into the mechanisms and dynamics underlying the control of spatio-temporal chaos in heterogeneous excitable media and provide new research perspectives towards alternative, life-saving low-energy defibrillation techniques.**

Spatially extended non-equilibrium systems display spatio-temporal dynamics that can range from ordered to turbulent. Controlling such systems is one of the central problems in nonlinear science and has far-reaching technological consequences. Few examples of successful control with applications in physics and chemistry have been demonstrated<sup>8,9</sup>. In biological excitable media, the systems' complexity makes successful control challenging. This difficulty applies in particular to electrical turbulence in cardiac tissue, known as fibrillation. During fibrillation, synchronous contraction of the muscle is disrupted by fast, vortex-like, rotating waves of electrical activity<sup>1–4</sup>. At the core of the vortex is a line of phase singularities called a filament. It is known that vortex instabilities and interactions<sup>10,11</sup> lead to self-organized, turbulent electrical dynamics. In the heart, electric turbulence arises in electro-mechanically anisotropic and heterogeneous cardiac muscle, which has complex geometry<sup>12</sup>. As we show below, this natural complexity can be used as a substrate for successful control of electrical turbulence.

The physiological mechanisms underlying the dynamics and control of electrical turbulence remain largely unknown<sup>13</sup>. The only clinically effective method for eliminating vortices in the heart is the delivery of a high-energy electric shock that both depolarizes and hyperpolarizes the tissue with a voltage gradient of about  $5 \text{ V cm}^{-1}$ . When applied externally, this shock can be as large as  $360 \text{ J}$  ( $1 \text{ kV}$ ,  $30 \text{ A}$ ,

$12 \text{ ms}$ )<sup>5</sup>. Although defibrillators that use this approach are used routinely in emergency medicine, treatments are often associated with severe side effects<sup>6,7,14</sup>.

Here we provide a new understanding of the biophysical mechanisms involved in the control of cardiac fibrillation and demonstrate low-energy control and termination of cardiac fibrillation *in vivo*. Two internal catheters with coiled wire electrodes were inserted into the right and left atria of adult beagle dogs (Fig. 1a) and sustained atrial fibrillation (AF) was induced (see Methods). We compared the energies required for defibrillation using a single, high-energy shock (standard defibrillation) and a sequence of five low-energy electric field pulses (low-energy antifibrillation pacing (LEAP)). A representative time series for successful LEAP termination via a monophasic action-potential electrode inserted into the right atrium is shown in Fig. 1b. For  $t < 0$ , the signal shows irregular oscillations with a dominant frequency  $f_v$  of  $6.8 \pm 0.1 \text{ Hz}$ . The control interval starts at  $t = 0$  (Fig. 1b, grey shaded area) and after control, the arrhythmia is terminated and normal sinus rhythm is restored. In this example, the energy required for terminating AF was  $0.074 \pm 0.012 \text{ J}$ , seven times less than the energy needed for standard defibrillation in this preparation ( $0.52 \pm 0.20 \text{ J}$ ). This substantial reduction was reproduced in 56 episodes in seven *in vivo* experiments, in which we found an average energy reduction of 84% (see Fig. 1d) ( $P < 10^{-7}$ ).

To identify the biophysical mechanisms underlying LEAP, we conducted *in vitro* experiments with isolated, perfused atria, using the same electrode configuration as in the corresponding *in vivo* experiments. *In vitro*, fluorescence imaging allowed quantitative measurements of the propagation of action potentials on the surface of the tissue, with high spatial and temporal resolution (see Methods). To assess differences in LEAP effectiveness between the *in vivo* and *in vitro* preparations, three of the five *in vitro* preparations were derived from the hearts used in the *in vivo* experiments. A time series of the fluorescence signal is shown in Fig. 1c (same heart as in Fig. 1b). For  $t < 0$ , we observed sustained AF with a dominant frequency of  $f_v = 6.8 \pm 0.1 \text{ Hz}$ . After LEAP (Fig. 1c, grey shaded area), normal sinus rhythm was restored. The LEAP pulse energy was  $0.066 \pm 0.017 \text{ J}$ , compared to  $1.15 \pm 0.29 \text{ J}$  for standard, single-pulse defibrillation. The overall energy reduction in the *in vitro* experiments ( $n = 5$  preparations, 39 defibrillation episodes and 46 LEAP episodes) was 91% (Fig. 1d) ( $P < 10^{-7}$ ). There was no significant difference in the energy required *in vivo* and *in vitro* for LEAP ( $P = 0.49$ ) and conventional defibrillation ( $P = 0.63$ ). Our findings are also in agreement with a separate set of *in vitro* experiments<sup>15</sup> in canine right-atrial preparations ( $n = 8$ ), in which LEAP terminated AF with a success rate of 93%, using only 13% of the energy per pulse required for a single shock

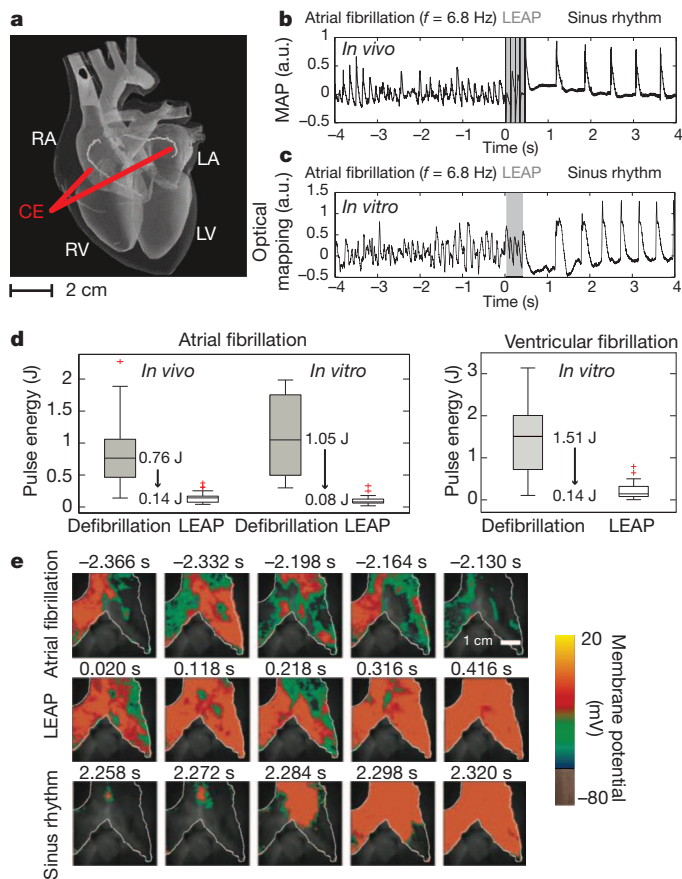
<sup>1</sup>Max Planck Institute for Dynamics and Self-Organization, Am Fassberg 17, D-37077 Göttingen, Germany. <sup>2</sup>Department of Biomedical Sciences, Cornell University, Ithaca, New York 14853, USA. <sup>3</sup>Institute for Nonlinear Dynamics, Georg August University, Am Fassberg 17, D-37077 Göttingen, Germany. <sup>4</sup>Heart Research Center Göttingen (HRCG), Robert-Koch-Strasse 40, D-37075 Göttingen, Germany.

<sup>5</sup>Department of Clinical Sciences, Cornell University, Ithaca, New York 14853, USA. <sup>6</sup>Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA. <sup>7</sup>Department of Cardiology and Pneumology, University Medical Center (UMG), Robert-Koch-Strasse 40, D-37075 Göttingen, Germany. <sup>8</sup>School of Mathematical Sciences, Rochester Institute of Technology, Rochester, New York 14623, USA. <sup>9</sup>Institut Non-Linéaire de Nice, F-06560 Valbonne, France. <sup>10</sup>Laboratoire de Physique de l'Ecole Normale Supérieure de Lyon, Université Lyon 1 and CNRS, F-69007 Lyon, France.

<sup>11</sup>Department of Mechanical and Aerospace Engineering, Cornell University, Ithaca, New York 14853, USA.

\*These authors contributed equally to this work.





**Figure 1 | Low-energy termination of cardiac electrical turbulence *in vivo* and *in vitro*.** **a**, Schematic of the anatomy of the heart. LA, left atrium; LV, left ventricle; RA, right atrium; RV, right ventricle. A pulsed electric field was applied with standard cardioversion coiled wire electrodes (CE) inserted into the left and right atria by catheters (see Supplementary Information). **b**, Monophasic action potential (MAP) recording of termination of AF using LEAP *in vivo*. a.u., arbitrary units. Dominant frequency  $f_v = 6.8 \pm 0.1$  Hz,  $n = 5$  pulses, pulse duration  $\Delta t = 8$  ms, pacing cycle length  $T_p = 99$  ms, pulse energy  $W = 0.074 \pm 0.012$  J. **c**, Termination of AF *in vitro*, measured from the atrial epicardium of the same heart as in **b**, by optical mapping (see **e**). The signal from a  $0.3 \times 0.3$  mm<sup>2</sup> region is shown ( $f_v = 6.8 \pm 0.1$  Hz,  $n = 5$ ,  $\Delta t = 8$  ms,  $T_p = 90$  ms,  $W = 0.066 \pm 0.017$  J). **d**, Reduction in pulse energy using LEAP versus standard defibrillation. *In vivo* AF ( $n = 7$ ): LEAP (56 episodes, mean energy  $\bar{W} = 0.14 \pm 0.08$  J); defibrillation (22 episodes,  $\bar{W} = 0.89 \pm 0.56$  J). *In vitro* AF ( $n = 5$ ): LEAP (46 episodes,  $\bar{W} = 0.10 \pm 0.07$  J); defibrillation (39 episodes,  $\bar{W} = 1.15 \pm 0.58$  J). *In vitro* ventricular fibrillation ( $n = 7$ ): LEAP (28 episodes,  $\bar{W} = 0.17 \pm 0.16$  J); defibrillation (12 episodes,  $\bar{W} = 1.34 \pm 0.89$  J; see Supplementary Information). Box plots show the median and the 25th and 75th percentiles. Whiskers indicate the statistically significant data range and red crosses mark outliers. **e**, Optical mapping of the AF termination that is also shown in **c**. During AF, complex spatio-temporal propagation of electrical excitation waves was observed (white line indicates boundary of atrium). LEAP ( $n = 5$ ,  $\Delta t = 90$  ms) progressively synchronized the tissue (see Supplementary Movies 1 and 2). Data are given as mean  $\pm$  s.d. unless stated otherwise.

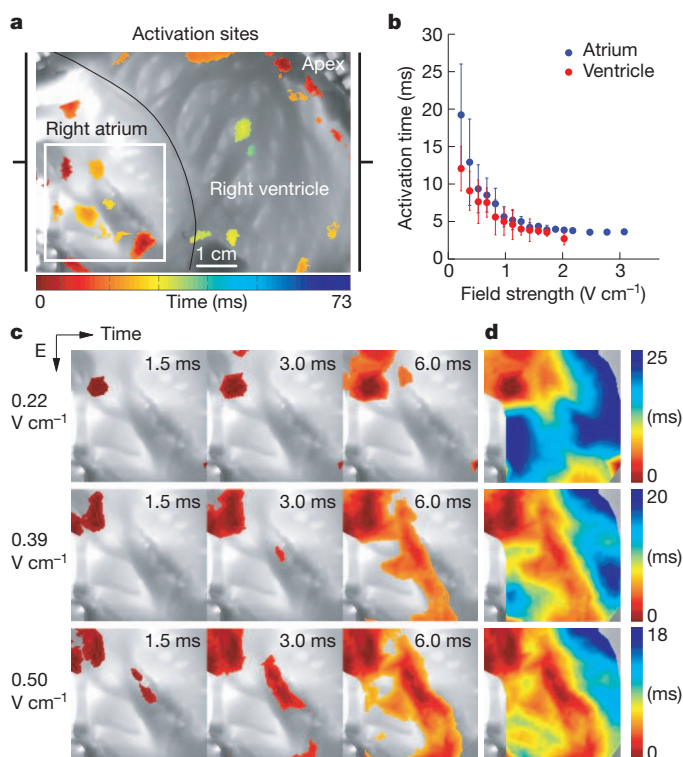
( $P < 0.002$ ). Furthermore, LEAP effectiveness was demonstrated for ventricular fibrillation *in vitro* ( $n = 7$  canine preparations, 12 defibrillation episodes and 28 LEAP episodes). In these experiments (Fig. 1d), the average energy reduction of LEAP versus a single shock was 85% ( $P < 10^{-2}$ ). The spatio-temporal excitation dynamics of the right atrium *in vitro* before, during and after control are shown in Fig. 1d (see also Supplementary Movie 1). During fibrillation, waves of turbulent electric activation propagate across the atria. At  $t = 0$ , a sequence of five electric pulses is applied at the coil electrodes, followed by a transient, spatio-temporal reorganization of the activation waves. After each pulse, the area that is activated increases, indicating progressive synchronization of the myocardium; fibrillation then

terminates and normal sinus rhythm (Supplementary Movie 2) can resume.

To elucidate the mechanism of defibrillation by LEAP, we studied the response of quiescent atrial and ventricular tissue to a homogeneous, pulsed electric field (Fig. 2a, b). In Fig. 2c, images taken at 1.5 ms, 3 ms and 6 ms after the pulse ( $0.22$  V cm<sup>-1</sup>) show depolarization induced by a single source. However, with increasing electric field strengths of  $0.22$  V cm<sup>-1</sup>,  $0.39$  V cm<sup>-1</sup> and  $0.50$  V cm<sup>-1</sup>, the number of sources increases to several dozen over the entire tissue. The locations of these sources and the wave propagation patterns are summarized in the isochronal maps shown in Fig. 2d. The density of sources, shown in Fig. 2c and d, increased with increasing field strength for both the ventricle and the atrium, thereby decreasing the activation time (Fig. 2b).

The results can be explained in the context of virtual electrodes<sup>15–20</sup>. In the bidomain representation, the voltage in cardiac tissue is the potential drop between the intracellular and extracellular medium. Theory predicts<sup>16</sup> that, in the presence of an electric field, discontinuities in tissue conductivity, such as blood vessels, changes in fibre direction, fatty tissue and intercellular clefts, induce a redistribution of intracellular and extracellular currents that can locally hyperpolarize or depolarize the cells. At the depolarization threshold, an excitation wave is emitted<sup>15–17,21</sup>.

The electric field that is necessary to produce an activation, as a function of the size of the conduction discontinuity in quiescent tissue,



**Figure 2 | Sites of activation in a cardiac preparation.** **a**, Canine wedge preparation ( $7.5 \times 5.6$  cm<sup>2</sup>) consisting of right atrium and right ventricle. At  $t = 0$  s, an electric field pulse of strength  $E = 0.34$  V cm<sup>-1</sup> was applied for 5 ms. The colour indicates the time of local activation observed with fluorescence imaging on the endocardium; the greyscale trans-illumination image shows the complex anatomy of the endocardium. The white square marks the area shown in panel **c**. **b**, Mean activation times  $\tau(E)$  for atria (blue circles,  $n = 3$  preparations, 17 measurements of  $\tau(E)$ ) and ventricles (red circles,  $n = 6$  preparations, 24 measurements of  $\tau(E)$ ) in response to an electric field pulse of strength  $E$  and duration 5 ms. Error bars indicate s.d. **c**, Activation of the atrium (in the region indicated by the white square in panel **a**) after an electric field pulse at  $t = 0$ . With increasing field strength, the number of activation sites increased and the time interval for total activation decreased. The colour code for each row is shown in **d** (see Supplementary Movie 3). For  $E < 0.2$  V cm<sup>-1</sup>, no waves were observed. **d**, Isochronal maps of the activation sequences shown in **c**.



can be estimated by approximating the discontinuous geometries (circles in two dimensions, spheres in three dimensions) and linearizing the bidomain model equations<sup>16,21,22</sup> around the resting membrane potential<sup>21</sup>, obtaining

$$\nabla^2 e - \frac{e}{\lambda^2} = 0 \quad (1)$$

where  $e = \Phi - \Phi_{\text{rest}}$ ,  $\Phi$  and  $\Phi_{\text{rest}}$  are the induced and resting membrane potentials and  $\lambda$  ( $\sim 0.35$  mm) is the space constant. For a non-conducting tissue region of radius  $R$ , the minimum electric field,  $E$ , necessary to bring the voltage above the excitation threshold  $\Phi_t$  at the boundary ( $r = R$ ) is given by

$$E = -\frac{\Phi_t - \Phi_{\text{rest}}}{\lambda} \frac{K_1'(R/\lambda)}{K_1(R/\lambda)} \quad (2)$$

where  $K_1(R/\lambda)$  is the modified Bessel function of the second kind (see Supplementary Information). Qualitatively, equation (2) indicates that the smaller the size of the heterogeneity, the larger the electric field ( $E$ ) required to emit waves. Equivalently, when  $E$  increases, discontinuities with a size larger than  $r \geq R_{\text{min}}(E)$  are recruited as wave sources, where  $R_{\text{min}}(E)$  is obtained by solving equation (2).  $R_{\text{min}}(E) \approx 1/E$  when  $E$  is large (Supplementary Fig. 13). The heart has heterogeneities of all sizes  $R$ , with a distribution  $p(R)$ . Assuming that these heterogeneities are uniformly distributed in the tissue, the density of recruited wave sources is

$$\rho(E) = \frac{N_{\text{total}}}{V} \int_{R_{\text{min}}(E)}^{R_{\text{max}}} p(R) dR \quad (3)$$

where  $R_{\text{max}}$  is the size of the largest discontinuity in the tissue.

To quantify  $p(R)$  associated with blood vessels in intact cardiac tissue, the coronary arteries of eight cardiac preparations were perfused with contrast agent and scanned using micro-computed tomography (micro-CT; Supplementary Figs 1–3 and 12). As shown in Fig. 3a and e, the size distribution  $p(R)$  of discontinuities yielded power laws  $p(R) \propto R^\alpha$  with exponents  $\alpha = -2.74 \pm 0.05$  ( $n = 5$  preparations, Supplementary Figs 4 and 6 and Supplementary Table 1) for atria and  $\alpha = -2.75 \pm 0.30$  ( $n = 3$  preparations, Supplementary Figs 4 and 5 and Supplementary Table 2) for ventricles. In biological systems, power-law scaling reflects generic underlying physical and physiological design principles relating form and function<sup>23,24</sup>.

The geometric structure of the coronary vasculature results in characteristic activation dynamics in response to a pulsed electric field, so the activation times as a function of field strength that are shown in Fig. 2b can be predicted using equation (3) and  $p(R)$ . The excitation wave emitted by a single heterogeneity and propagating radially at constant velocity  $v$  excites a volume  $V(\tau) = 4\pi(v\tau)^3/3$  in a time interval  $\tau$ . For  $N$  heterogeneities, uniformly distributed with density  $\rho = N/V$ , the entire tissue is excited after

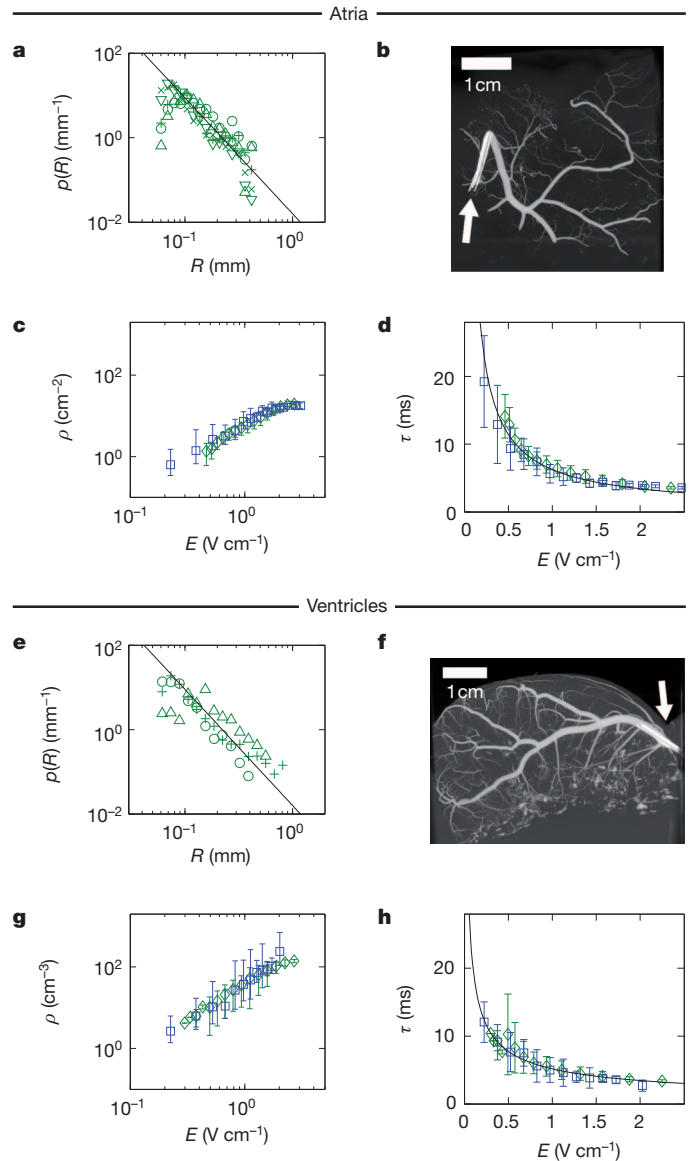
$$\tau(\rho) = (3/(4\pi\rho))^{1/3}/v \quad (4)$$

Equations (2) to (4) quantitatively relate structural properties (that is, the size distribution) to functional dynamics (that is, the activation times). In particular, these equations indicate (see Supplementary Information) that the exponents of size distributions,  $p(R) \propto R^\alpha$ , and activation times,  $\tau \propto E^\beta$ , are related, for large  $E$ , by

$$\beta = (1 + \alpha)/d \quad (5)$$

where  $d$  is the geometric dimension of the tissue. To assess the role of tissue geometry, we measured the scaling exponents  $\alpha$  and  $\beta$  for the thick ventricular wall ( $d = 3$ ) and the thin, quasi-two-dimensional atrial wall ( $d = 2$ ). We used the measured size distribution exponents,  $\alpha$ , to estimate the activation time exponents,  $\beta_p$ , using equation (5). The predicted exponents  $\beta_p$  were found to be in good quantitative agreement with the observed exponents  $\beta_o$  for atrial and ventricular tissue (Table 1, Supplementary Figs 7–9 and Supplementary Tables 3 and 4).

From the measured size distributions  $p(R)$  shown in Fig. 3a, we numerically estimated the intramural wave source densities  $\rho(E)$  using



**Figure 3 | From anatomical structure to activation dynamics in atria (a–d) and ventricles (e–h).** **a**, Probability distribution of radii,  $p(R)$ , of canine coronary arteries in atrial tissue, obtained from micro-CT measurements (symbols represent  $n = 5$  different preparations, for details see Supplementary Information). The black line indicates the power law  $p \propto R^{-2.74 \pm 0.05}$  (mean scaling exponent of all preparations). **b**, **f**, Examples of atrial (**b**) and ventricular (**f**) anatomical structure of coronary arteries (see Supplementary Movies 9 and 10). White arrows indicate the position of the catheters used to inject the contrast agent (see Methods). **c**, Density of wave sources derived from equation (3) and from the atrial measurements shown in panel **a** (green diamonds), and corresponding density estimated from the activation-time measurements shown in panel **d** (blue squares). The predicted density from the structural data in **a** is plotted as the mean of the predictions from individual preparations. **d**, Atrial activation-time measurements using optical mapping (blue squares), and corresponding prediction of activation dynamics (green diamonds), on the basis of the source density obtained in **c** from the size distribution in **a** (plotted as the mean of predictions from individual preparations). The black line indicates the power law  $\tau \propto E^{-0.87 \pm 0.03}$  (see Table 1 and Supplementary Information). **e**, Probability distribution  $p(R)$  of coronary artery radii for ventricular tissue ( $n = 3$ ). The black line indicates the power law  $p \propto R^{-2.75 \pm 0.3}$  (mean scaling exponents of all preparations). **g**, Density of wave sources derived from the ventricular measurements shown in **e** (green diamonds), and corresponding density estimated from activation-time measurements shown in **h** (blue squares). **h**, Ventricular activation-time measurements (blue squares) and prediction of activation times (green diamonds) based on  $p(R)$ . The black line indicates the power law  $\tau \propto E^{-0.58 \pm 0.10}$  (see Table 1 and Supplementary Information). Error bars indicate s.d. in all plots. Additional three-dimensional illustrations and animations are available at <http://thevirtualheart.org/vessels>.

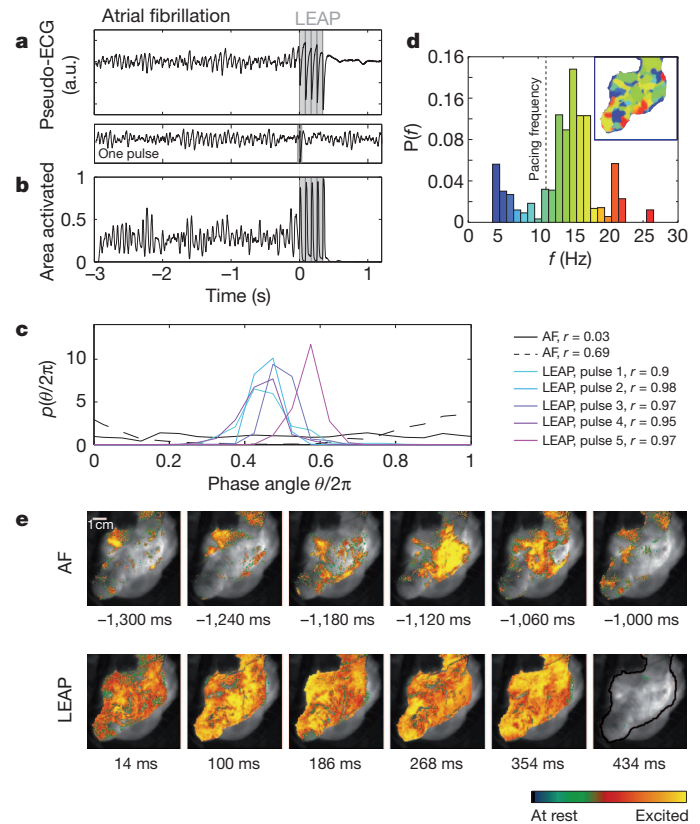
equation (3) (green symbols in Fig. 3c), and  $\tau(E)$  using equation (4) (green symbols in Fig. 3d, see Supplementary Information). Similarly, from the measured activation time  $\tau(E)$  shown in Fig. 3d (blue symbols), we numerically estimated the intramural wave source density  $\rho(E)$  (blue symbols in Fig. 3c) using the inverse of equation (4). Figure 3c, d, g and h shows comparisons between measured and calculated data for atria and ventricles, respectively. Green symbols represent results based on the measured size distribution  $p(R)$ , whereas results based on the measured activation time  $\tau(E)$  are shown in blue. Again, excellent agreement was found (Table 1). Our results show that the structural properties of the coronary vasculature quantitatively describe the timescales of tissue activation.

The experimental results shown in Fig. 3 indicate that, in addition to the known effect of recruiting tissue boundaries during an electric shock<sup>25</sup>, the heterogeneity of the vascular structure also allows the recruitment of an effective density of wave sources inside the myocardium. These distributed wave sources can be used for non-invasive, intramural multi-site pacing, which is key to the novel approach of LEAP, as demonstrated in Fig. 1. Previous studies aimed to control a single, isolated vortex attached to a large heterogeneity<sup>21,26,27</sup> and it was thought that in tissue with many heterogeneities of different sizes, it should be possible to control disordered regimes<sup>28</sup> by varying the electric field strength to control the number of wave sources<sup>29</sup>. An experimental demonstration of this approach was provided for AF<sup>15</sup>; however, the underlying control mechanism was not explored.

As shown in Fig. 3, strengthening the electric field increases the density of wave sources. Thus, the probability of wave sources, both in regions of excitable tissue and in the vicinity of filaments, increases. Being excitable, these regions can be brought above the excitation threshold and fully depolarized by the applied electric field pulse. Thus, they act as control sites that directly affect vortex filaments, which are the source of fibrillation.

Figure 4 provides experimental evidence that LEAP interacts directly with multiple vortices simultaneously. During AF, complex spatio-temporal dynamics with several interacting waves are observed (Fig. 4e and Supplementary Movie 10), evidenced by the presence of multiple phase singularities. During the episode of AF shown in Fig. 4, we observed  $9.9 \pm 4.4$  (mean  $\pm$  s.d.) phase singularities, resulting in the spatial complexity of the dominant frequency map (Fig. 4d, inset) and the corresponding broad probability distribution of frequencies, centred around 15 Hz (Fig. 4d). The observed spatio-temporal complexity can also be found in the pseudo-electrocardiogram (pseudo-ECG) and the percentage of area activated (PAA) (Fig. 4a, b and Supplementary Information). Whereas both signals show complex amplitude fluctuations during AF, the periodic perturbations during LEAP defibrillation result in increasingly coherent dynamics in the entire tissue, associated with a progressive increase in amplitude of the pseudo-ECG and the PAA. During AF, the normalized PAA is  $0.28 \pm 0.11$  (mean  $\pm$  s.d.), whereas during LEAP, PAA increases with each pulse towards one, that is, simultaneous activation of the entire tissue. After LEAP, AF is terminated and periodic, normal rhythm resumes (Supplementary Movie 10).

Figure 1 demonstrates the efficacy of LEAP defibrillation and Fig. 4 shows that tissue synchronization, and therefore control of fibrillation, is achieved by perturbing the system near the filaments, where it is most susceptible to perturbations<sup>30</sup>. A periodic perturbation can reach a nearby filament only if  $f_{\text{LEAP}} > f_v$ , where  $f_{\text{LEAP}}$  is the frequency of the



**Figure 4 | Direct access to vortex cores.** **a**, Termination of fibrillation with LEAP. The pseudo-ECG is obtained from the optical mapping experiments (see Supplementary Information). The dominant frequency during AF is  $f_v = 15.0 \pm 0.5$  Hz (see inset in **d**). The vertical grey lines indicate the times at which five LEAP pulses ( $f_{\text{LEAP}} = 11.8$  Hz,  $E = 1.96$  V cm<sup>-1</sup>) were delivered. A single pulse at this field strength did not terminate fibrillation (see pseudo-ECG below panel **a**). After LEAP, normal rhythm was resumed. The termination of fibrillation with  $f_{\text{LEAP}} < f_v$  was observed in atria ( $n = 4$  atria, 6 episodes) and ventricles ( $n = 1$  ventricle, 3 episodes). **b**, The area activated indicates tissue synchronization during LEAP. Grey lines indicate timing of LEAP pulses. **c**, Probability distribution of phases  $\theta_j$  during AF, and for the times of strongest synchronization (maximum order parameter  $r$ , see legend) after each LEAP pulse (see Supplementary Information). During AF, broad phase distributions indicate partial coherence (solid and dashed black lines show two representative distributions). LEAP with  $f_{\text{LEAP}} < f_v$  induces synchronization and thus termination of AF. **d**, Probability distribution of dominant frequencies during AF, obtained from optical mapping. The dominant frequency map (inset) shows a complex spatial domain structure corresponding to multiple interacting waves (for colour code see histogram;  $f_v = 15.0 \pm 1.0$  Hz;  $f_{\text{LEAP}} = 11.8 \pm 0.5$  Hz, indicated by a vertical dashed line). **e**, Spatio-temporal dynamics during AF, LEAP (images taken at the times of maximum-area-activated after each pulse; see **b**) and sinus rhythm (see Supplementary Movie 8). The greyscale image shows the atrium (see black line in the last image;  $70 \times 70$  mm<sup>2</sup>). The last image shows quiescence after AF termination.

perturbation and  $f_v$  is the frequency of the associated vortex. However, if  $f_v$  exceeds  $f_{\text{LEAP}}$ , a distant wave source cannot perturb the filament, owing to wave annihilation (Supplementary Figs 10 and 11). Nevertheless, as shown in Fig. 4a and b, AF was terminated with  $f_v > f_{\text{LEAP}}$ . We quantified termination of multiple filaments with  $f_v > f_{\text{LEAP}}$  in atrial tissue ( $n = 4$  preparations, 6 episodes) and ventricular tissue ( $n = 1$  preparation, 3 episodes) (Supplementary Table 5). This result demonstrates simultaneous, local access to multiple filaments. Our experiments indicate that by adjusting the number of recruited sites, this approach may successfully terminate fibrillation regardless of whether fibrillation is caused by multiple re-entrant waves or a single mother rotor<sup>1-3</sup>. Consequently, this mechanism is applicable to multiple underlying causes of fibrillation, in both atria and ventricles. Our findings on the mechanism of defibrillation,

**Table 1 | Observed and predicted scaling exponents**

Tissue	$d$	$\alpha_0$	$\beta_0$	$\beta_p^\dagger$	$\beta_p^*$
Atrium	2	$-2.74 \pm 0.05$	$-0.81 \pm 0.23$	$-0.88 \pm 0.20$	$-0.87 \pm 0.03$
Ventricle	3	$-2.75 \pm 0.30$	$-0.75 \pm 0.18$	$-0.74 \pm 0.25$	$-0.58 \pm 0.10$

$p(R)$  and the activation time  $\tau(E)$  for atria and ventricles. Statistical analysis showed no significant difference for  $\alpha_0$  for atria and ventricles (see Supplementary Information).  $\beta_p^*$  as obtained from the high-field-strength approximation equation (5).  $\dagger$  Average of  $\beta_p$  obtained from least squares fits to activation times, obtained from direct numerical estimation using  $p(R)$  and equations (2) to (4) (Supplementary Tables 1–4).

together with the *in vivo* proof of defibrillation of atria with LEAP, should allow the development of new approaches towards alternative life-saving defibrillation techniques.

## METHODS SUMMARY

Experiments were conducted in open-chest anaesthetized dogs or in isolated, arterially perfused canine atrial and ventricular preparations *in vitro* (see Methods). Atrial fibrillation was induced *in vivo* using a pace-down protocol during stimulation of the right vagus nerve (2–4 mA at 10 Hz), and was induced *in vitro* during perfusion with acetylcholine (1–3  $\mu$ M). Ventricular fibrillation was induced *in vitro* using rapid pacing. Membrane potential was recorded *in vitro* by optical mapping using a voltage-sensitive AminoNaphthylEthenylPyridinium dye (di-4-ANEPPS). Standard 6.5F cardioversion catheters were used to deliver defibrillation shocks *in vivo* and *in vitro*. The shocks consisted of 1–5 symmetrical biphasic pulses of 8-ms duration at shock strengths of 20–100 V, delivered via a custom-built cardioverter/defibrillator. Immediately after the optical mapping experiments, tissues were injected with 1–2 ml of Microfil contrast agent at 0.05–0.15 ml min<sup>-1</sup> via the same cannula used for perfusion. The chambers were then filled with silicone to preserve tissue morphology during scans performed using a GE 120 micro-CT scanner with 25- $\mu$ m *x*–*y*–*z* resolution, to determine blood vessel sizes and distributions.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 9 September 2010; accepted 17 May 2011.

- Davidenko, J. M., Pertsov, A. V., Salomonsz, R., Baxter, W. & Jalife, J. Stationary and drifting spiral waves of excitation in isolated cardiac muscle. *Nature* **355**, 349–351 (1992).
- Gray, R. A., Pertsov, A. M. & Jalife, J. Spatial and temporal organization during cardiac fibrillation. *Nature* **392**, 75–78 (1998).
- Witkowski, F. X. *et al.* Spatiotemporal evolution of ventricular fibrillation. *Nature* **392**, 78–82 (1998).
- Cherry, E. M. & Fenton, F. H. Visualization of spiral and scroll waves in simulated and experimental cardiac tissue. *New J. Phys.* **10**, 125016–125059 (2008).
- Koster, R. W. *et al.* A randomized trial comparing monophasic and biphasic waveform shocks for external cardioversion of atrial fibrillation. *Am. Heart J.* **147**, e1–e7 (2004).
- Babbs, C. F., Tacker, W. A., VanVleet, J. F., Bourland, J. D. & Geddes, L. A. Therapeutic indices for transthoracic defibrillator shocks: effective, damaging, and lethal electrical doses. *Am. Heart J.* **99**, 734–738 (1980).
- Santini, M. *et al.* Single shock endocavitary low energy intracardiac cardioversion of chronic atrial fibrillation. *J. Interv. Card. Electrophysiol.* **3**, 45–51 (1999).
- Sakurai, T., Mihaliuk, E., Chirila, F. & Showalter, K. Design and control of wave propagation in excitable media. *Science* **296**, 2009–2012 (2002).
- Rappel, W. J., Fenton, F. H. & Karma, A. Spatiotemporal control of wave instabilities in cardiac tissue. *Phys. Rev. Lett.* **83**, 456–459 (1999).
- Biktashev, V. N., Holden, A. V. & Zhang, H. Tension of organizing filaments of scroll waves. *Phil. Trans. R. Soc. Lond. A* **347**, 611–630 (1994).
- Fenton, F. H., Cherry, E. M., Hastings, H. M. & Evans, S. J. Multiple mechanisms of spiral wave breakup in a model of cardiac electrical activity. *Chaos* **12**, 852–892 (2002).
- Fenton, F. H. & Karma, A. Vortex dynamics in three-dimensional continuous myocardium with fiber rotation: Filament instability and fibrillation. *Chaos* **8**, 20–47 (1998).
- Mackenzie, D. Making sense of a heart gone wild. *Science* **303**, 786–787 (2004).
- Walcott, G. P., Killingsworth, C. R. & Ideker, R. E. Do clinically relevant transthoracic defibrillation energies cause myocardial damage and dysfunction? *Resuscitation* **59**, 59–70 (2003).
- Fenton, F. H. *et al.* Termination of atrial fibrillation using pulsed low-energy far-field stimulation. *Circulation* **120**, 467–476 (2009).
- Plonsey, R. The nature of sources of bioelectric and biomagnetic fields. *Biophys. J.* **39**, 309–312 (1982).
- Fast, V. G., Rohr, S., Gillis, A. M. & Kléber, A. G. Activation of cardiac tissue by extracellular electrical shocks: Formation of ‘secondary sources’ at intercellular clefts in monolayers of cultured myocytes. *Circ. Res.* **82**, 375–385 (1998).
- Sambelashvili, A. T., Nikolski, V. P. & Efimov, I. R. Virtual electrode theory explains pacing threshold increase caused by cardiac tissue damage. *Am. J. Physiol. Heart Circ. Physiol.* **286**, H2183–H2194 (2004).
- Hooks, D. A. *et al.* Cardiac microstructure: Implications for electrical propagation and defibrillation in the heart. *Circ. Res.* **91**, 331–338 (2002).
- Trayanova, N. & Skouibine, K. Modeling defibrillation: Effects of fiber curvature. *J. Electrocardiol.* **31** (suppl.), 23–29 (1998).
- Pumir, A. & Krinsky, V. Unpinning of a rotating wave in cardiac muscle by an electric field. *J. Theor. Biol.* **199**, 311–319 (1999).
- Roth, B. J. & Wikswo, J. P. A bi-domain model for the extracellular potential and magnetic field of cardiac tissue. *IEEE Trans. Biomed. Eng.* **33**, 467–469 (1986).
- Murray, C. D. The physiological principle of minimum work: I. The vascular system and the cost of blood volume. *Proc. Natl Acad. Sci. USA* **12**, 207–214 (1926).
- Kassab, G. S. Scaling laws of vascular trees: of form and function. *Am. J. Physiol. Heart Circ. Physiol.* **290**, H894–H903 (2006).
- Maleckar, M. M. *et al.* Polarity reversal lowers activation time during diastolic field stimulation of the rabbit ventricles: insights into mechanisms. *Am. J. Physiol. Heart Circ. Physiol.* **295**, H1626–H1633 (2008).
- Ripplinger, C. M., Krinsky, V. I., Nikolski, V. P. & Efimov, I. R. Mechanisms of unpinning and termination of ventricular tachycardia. *Am. J. Physiol. Heart Circ. Physiol.* **291**, H184–H192 (2006).
- Takagi, S. *et al.* Unpinning and removal of a rotating wave in cardiac muscle. *Phys. Rev. Lett.* **93**, 058101 (2004).
- Kirchhof, C. *et al.* Regional entrainment of atrial fibrillation studied by high-resolution mapping in open-chest dogs. *Circulation* **88**, 736–749 (1993).
- Pumir, A. *et al.* Wave emission from heterogeneities opens a way to controlling chaos in the heart. *Phys. Rev. Lett.* **99**, 208101 (2007).
- Gray, R. & Chattipakorn, N. Termination of spiral waves during cardiac fibrillation via shock-induced phase resetting. *Proc. Natl Acad. Sci. USA* **102**, 4672–4677 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. L. Riccio and the Cornell University  $\mu$ CT Facility for Imaging and Preclinical Research for performing the micro-CT scanning, and T. K. Hitchens for technical assistance with cardiac structural imaging. M. W. Enyeart and J. Boesch assisted with the experiments and N. F. Otani provided insights from computer simulations. G. Hooker conducted the statistical analysis. This work was supported by National Science Foundation (NSF) grants 0800793 (F.H.F. and E.M.C.) and 0926190 (F.H.F.); by National Institutes of Health (NIH) grants HL075515-S04 (F.H.F.), HL075515 (R.F.G.) and HL073644 (R.F.G. and E.B.); by IFCPAR Project no. 3404-4 (A.P.); by the German Ministry for Education and Research through FKZ 01EZ0905/6 (S.L., M.Z., E.B. and G.H.); by the Kavli Institute for Theoretical Physics through NSFPHY05-51164; by the Pittsburgh Supercomputing Center (NSF TeraGrid); by the Pittsburgh NMR Center for Biomedical Research (NIH P41-EB001977); by the European Community's Seventh Framework Programme FP7/2007–2013 agreement HEALTH-F2-2009-241526 (EUTrigTreat) and by the Max Planck Society.

**Author Contributions** E.B., F.H.F., R.F.G., V.I.K., S.L. and A.P. designed research and wrote the paper with P.B., E.M.C., G.H., D.H., B.G.K. and A.S. P.B., F.H.F., S.L. and A.S. did the *in vitro* experiments and analysed the data. L.C., F.H.F., J.F., R.F.G., A.G., G.H., B.G.K., S.L., A.S. and M.Z. contributed to the *in vivo* experiments and F.H.F. and S.L. analysed the data. P.B., F.H.F. and A.S. did the CT scanning; P.B., E.B., E.M.C., F.H.F., D.H., V.I.K., G.L., S.L., A.P. and A.S. analysed and interpreted the data.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.L. ([stefan.luther@ds.mpg.de](mailto:stefan.luther@ds.mpg.de)) and F.H.F. ([flavio.h.fenton@cornell.edu](mailto:flavio.h.fenton@cornell.edu)).



## METHODS

**Cardiac preparation *in vivo*.** Adult beagle dogs ( $n = 7$ ) were induced with fentanyl citrate ( $0.05 \text{ mg kg}^{-1}$ ) and propofol ( $0.025 \text{ mg kg}^{-1}$ ), given as two sequential intravenous boluses. After intravenous administration of atracurium ( $0.4 \text{ mg kg}^{-1}$ ), the dogs were intubated, artificially ventilated with oxygen and maintained under general anaesthesia with a continuous-rate infusion of fentanyl ( $0.004 \text{ mg kg}^{-1} \text{ h}^{-1}$ ). A median sternotomy was performed and the heart was suspended in a pericardial cradle. Normal body temperature was maintained with an inflatable heated body jacket and fluid losses were replaced with lactated Ringer's solution (at  $10 \text{ ml kg}^{-1} \text{ h}^{-1}$ , intravenously). Blood oxygen saturation and carbon dioxide levels were monitored continuously and maintained within normal limits by varying the tidal volume and/or respiratory rate.

8 F introducers were placed in both femoral veins to allow the passage of catheters that would provide programmed stimulation to induce AF and to suppress AF using LEAP, as well as to monitor atrial activation. The catheters were 6.5 F and 160–175 cm long, with a single coiled-wire monofilament cardioversion electrode 6 cm in length situated 1 cm from the tip of the catheter (Modified Rhythm and Impact catheters, Rhythm Technologies Inc.). The catheters also carried two sensing electrodes distal to the coiled wire electrode. One catheter was placed via a femoral vein introducer and advanced until its coil electrode was positioned in the left pulmonary artery. A second catheter was placed so that its coil electrode was positioned in the right atrium. Alternatively, one of the catheters was inserted into the left atrium via a puncture wound in the left atrial appendage. The LEAP stimulus was then applied across the two catheter coils. In four of the dogs, LEAP stimulation was also delivered via patch electrodes sutured to the right and left atrial appendages. The patch electrodes consisted of  $2\text{-cm} \times 2\text{-cm}$  pieces of stainless steel wire mesh, insulated with rubber membrane on the surface not in contact with the atrial tissue. Atrial sensing was performed using a monophasic action potential (MAP) catheter (7 F, EP Technologies), which was advanced via the femoral vein into the right atrium.

The following signals were recorded: lead II surface ECG, MAP from the right atrium, bipolar electrograms from the right and left atria, and arterial blood pressure. The recordings were acquired at a sampling frequency of 1,000 Hz and stored digitally using a data acquisition system (BioPac Systems, MP 150; software, AcqKnowledge 3.7.3).

To permit stimulation of the vagus nerve and thereby facilitate the induction and maintenance of AF, the right cervical vagus nerve was isolated, doubly ligated and cut. Bipolar iridium wire electrodes, insulated except at the tip, were inserted into the sheath of the nerve and connected to a WPI stimulator. Immediately before AF induction and continuously during AF, pulses of 2-ms duration were delivered at a current strength that produced the maximum reduction in sinus rate (typically 2–4 mA) at a stimulus frequency of 10 Hz. The region surrounding the cut end of the nerve was bathed in mineral oil to prevent desiccation and preserve intact function over the 5–7-h time course of the experiment. AF was induced using a pace-down protocol, in which a train of symmetrical biphasic LEAP pulses of 8-ms duration (4 ms up/4 ms down) was delivered at an intensity of 2.0 V, initially at a cycle length of 300 ms. The cycle length was subsequently shortened progressively until AF was induced.

After instrumentation of the dog, the experimental protocol was: (1) determine the threshold for activation of the atria using a single symmetrical biphasic LEAP stimulus pulse (4 ms up/4 ms down), delivered at a constant cycle length of 400 ms; (2) determine the impedances of the LEAP electrodes by delivering single biphasic pulses of 8-ms duration at voltages of 10, 20, 40, 60, 80 and 100 V and measuring the resulting voltage deflection at the sensing electrode; (3) determine the intensity and frequency of vagal stimulation required to reduce sinus heart rate maximally; (4) induce AF using the pacedown protocol described above and monitor for 2 min; (5) attempt to cardiovert using a single electric shock of 8-ms duration, starting with a shock strength of 40 V and increasing the voltage by increments of 5–10 V until cardioversion occurred, up to a maximum of 100 V; (6a) if cardioversion was successful, reinstate AF and attempt to suppress AF using five symmetrical biphasic LEAP pulses of 8-ms duration (4 ms up/4 ms down), delivered at a cycle length 5–10 ms shorter than the cycle length corresponding to the dominant frequency during AF, as determined using a Fast Fourier Transform (FFT) of the MAP recording, and at an initial voltage of 10 V, followed by increments of 5–10 V until AF was suppressed; (6b) if cardioversion was not successful, attempt to suppress AF using LEAP stimulation; (6c) if neither cardioversion nor LEAP stimulation was successful, turn off vagal stimulation and suppress AF using a single shock or LEAP stimulation; (7) repeat steps 4–6 for an additional 3–5 trials; (8) change the LEAP electrode configuration (for example, from catheters to patch electrodes) and repeat steps 1–7.

The experimental procedures were approved by the institutional animal care and use committee of the Center for Animal Resources and Education at Cornell University.

**Cardiac preparation *in vitro*.** Adult beagle dogs of either sex, age 1–4 years ( $n = 5$  for the atrial and  $n = 7$  for the ventricular experiments), were anaesthetized with Fatal-Plus ( $390 \text{ mg ml}^{-1}$  pentobarbital sodium, Vortex Pharmaceuticals;  $86 \text{ mg kg}^{-1}$  intravenously) and their hearts were rapidly excised. For the atrial experiments, after excision, the right and left coronary arteries were cannulated using polyethylene tubing and the right and left atria were excised and perfused with normal Tyrode solution bubbled with 95%  $\text{O}_2$ , 5%  $\text{CO}_2$  at  $p_{\text{O}_2}$  400–600 mm Hg,  $\text{pH} 7.35 \pm 0.05$  and temperature  $37.0 \pm 0.5^\circ\text{C}$ . The flow rates of the perfusate and superfusate were  $20 \text{ ml min}^{-1}$  and  $60 \text{ ml min}^{-1}$ , respectively, at a perfusion pressure of 50–80 mm Hg. After 15–30 min of equilibration, the preparation was stained with the voltage-sensitive dye di-4-ANEPPS ( $10 \mu\text{mol l}^{-1}$  bolus). Blebbistatin ( $10 \mu\text{mol l}^{-1}$  constant infusion over 30–40 min) was added to prevent motion artefacts.

LEAP pulses were delivered from a custom-built cardioverter/defibrillator, which was capable of generating pulse sequences with a specified number of pulses, pulse duration, polarity and shape. The computer-controlled device used a digital-to-analog converter (NI USB-6259 BNC) to generate arbitrary waveforms, which were amplified using a power amplifier (Kepco BOB 100-4M). The waveform was configured manually using a Labview program or was loaded from a database. Electrophysiological signals and various monitor signals (for example, delivered current and voltage) were recorded using an analog-to-digital converter (NI USB-6259 BNC). The signals were analysed in real time and used to select the LEAP parameters automatically. An automated impedance measurement was used to calibrate the pulse energy. The data obtained during the experiment were stored in a database for offline analysis.

Standard bipolar stimulating electrodes were placed on the right and left atria (in the same positions as in the corresponding *in vivo* experiments) and field stimulation was delivered as described above. To calculate the energy delivered, the impedance of the electrodes was determined by delivering single biphasic pulses of 8-ms duration at voltages of 10, 20, 40, 60, 80 and 100 V and measuring the resulting voltage deflection at the sensing electrode. Pacing stimuli were delivered using a WPI stimulator and stimulus isolator, and LEAP stimuli were delivered using a function generator and custom-built current source. Field strengths in excess of  $5 \text{ V cm}^{-1}$  could be delivered using this device, at cycle lengths as short as 50 ms. The field strength between the electrodes was measured using two teflon-coated silver wires immersed in the bath and set 5–10 mm apart.

To verify tissue viability, the excitation threshold for far-field stimulation in the quiescent myocardium was determined by monitoring optical wave activity after application of a sequence of five LEAP pulses, and was compared to the corresponding *in vivo* study values. AF was initiated subsequently using rapid pacing, either with or without acetylcholine ( $1\text{--}3 \mu\text{M}$ ) in the perfusate. The concentration of acetylcholine was titrated to induce AF with a similar dominant frequency to that observed *in vivo*. The presence or absence of AF was documented by monitoring wave activity optically. The dominant frequency during AF was determined from the FFT of a single pixel recording that could be moved in real time to assess the range of frequencies throughout the tissue. At the end of the study, the excitation threshold was recalculated to confirm the stability of the tissue. For the ventricular experiments, the protocols were similar to those used for the atrial experiments, as described in detail previously<sup>15</sup>. However, no acetylcholine was used and fibrillation was initiated by rapid pacing.

The experimental procedures were approved by the institutional animal care and use committee of the Center for Animal Resources and Education at Cornell University.

**Optical fluorescence imaging.** Illumination was provided by LEDs (Luxeon, 5 W, 530 nm). High-numerical-aperture lenses (F-number 0.95, focal length 50 mm) were fitted with long-wavelength-pass emission filters (580 nm). The epicardium and endocardium were imaged simultaneously using two synchronized cameras: the control-site studies used two high-resolution, high-speed CMOS cameras (Vision Research Phantom V7,  $600 \times 800$  pixels, 12-bit, 2,000 frames  $\text{s}^{-1}$ ), whereas arrhythmia termination experiments used two EMCCD cameras (electron multiplied charge coupled device, Photometrics Cascade 128+,  $128 \times 128$  pixels, 16-bit, 511 frames  $\text{s}^{-1}$ ). The pseudo-ECG is defined as the mean of the optical fluorescence signal over the entire field of view.

**Micro-computed tomography.** Immediately after optical mapping experiments, tissues were injected with 1–2 ml of Microfil contrast agent (Flow Tech Inc.) at  $0.05\text{--}0.15 \text{ ml min}^{-1}$  via the same cannula used for perfusing the tissue. The chambers were then filled with silicone to preserve tissue morphology during the scan. The contrast agent and silicone were allowed to set for at least 4 h before being scanned. The scans were performed using the GE CT120 micro-CT scanner (GE Healthcare). For each data set, 1,200 projections were obtained at  $0.3^\circ$  intervals over  $360^\circ$ , using 80 keV, 32 mA and  $25 \mu\text{m}$   $x\text{--}y\text{--}z$  resolution. Before each scan, ten bright-field images were acquired with no objects in the field of view, providing a correction for detector non-uniformity. Each image data set was transferred from the CT120

system to an image-processing workstation (HP xw8400 with 8 CPU cores and 16GB RAM). The projection views were used to reconstruct a CT image using a convolution back-projection approach implemented in three dimensions, giving a  $40 \times 40 \times 40 \text{ mm}^3$  volume of image data with  $25 \mu\text{m}$  or  $50 \mu\text{m}$  isotropic voxels in analog-to-digital units. Correction for signal non-uniformity across the field of view was determined from measurement within a water/tissue phantom (SB201), scanned with the same X-ray protocol. Ten axial slices within this water phantom were averaged to

produce a two-dimensional map of offset values, used to correct for non-uniformity at any point in the image. Corrected image data sets were calibrated to the conventional scale of Hounsfield units, defined so that water and air have values of 0 and  $-1,000$ , respectively. The maximum intensity projection (MIP) shown in Fig. 3b and f is an orthographic projection which maps voxels with maximum intensity that intersect parallel rays from the viewpoint to the plane of projection. The details of the further analysis are given in Supplementary Information.

# Architecture of the Mediator head module

Tsuyoshi Imasaki<sup>1</sup>, Guillermo Calero<sup>2†</sup>, Gang Cai<sup>3†</sup>, Kuang-Lei Tsai<sup>3</sup>, Kentaro Yamada<sup>1</sup>, Francesco Cardelli<sup>1</sup>, Hediye Erdjument-Bromage<sup>4</sup>, Paul Tempst<sup>4</sup>, Imre Berger<sup>5</sup>, Guy Lorch Kornberg<sup>2</sup>, Francisco J. Asturias<sup>3</sup>, Roger D. Kornberg<sup>2</sup> & Yuichiro Takagi<sup>1</sup>

**Mediator is a key regulator of eukaryotic transcription<sup>1</sup>, connecting activators and repressors bound to regulatory DNA elements with RNA polymerase II<sup>1–4</sup> (Pol II). In the yeast *Saccharomyces cerevisiae*, Mediator comprises 25 subunits with a total mass of more than one megadalton (refs 5, 6) and is organized into three modules, called head, middle/arm and tail<sup>7–9</sup>. Our understanding of Mediator assembly and its role in regulating transcription has been impeded so far by limited structural information. Here we report the crystal structure of the essential Mediator head module (seven subunits, with a mass of 223 kilodaltons) at a resolution of 4.3 Å. Our structure reveals three distinct domains, with the integrity of the complex centred on a bundle of ten helices from five different head subunits. An intricate pattern of interactions within this helical bundle ensures the stable assembly of the head subunits and provides the binding sites for general transcription factors and Pol II. Our structural and functional data suggest that the head module juxtaposes transcription factor IIH and the carboxy-terminal domain of the largest subunit of Pol II, thereby facilitating phosphorylation of the carboxy-terminal domain of Pol II. Our results reveal architectural principles underlying the role of Mediator in the regulation of gene expression.**

In the yeast *S. cerevisiae*, the Mediator head module is composed of seven subunits<sup>10</sup>: Med17 (also known as Srb4), Med11, Med22 (Srb6), Med6, Med8, Med18 (Srb5) and Med20 (Srb2). Four subunits are encoded by *SRB* genes, first identified through a genetic screen for mutations suppressing the Pol II carboxy-terminal domain (CTD) truncation<sup>11,12</sup>. The head module is essential for Mediator function because mutations in the head abolish messenger RNA synthesis *in vivo*<sup>13,14</sup> and *in vitro*<sup>15</sup>, and eliminate Mediator interaction with promoters *in vivo*<sup>10</sup>. The head module is organized into three domains that can undergo significant conformational changes, and it interacts with the TATA-binding protein subunit of general transcription factor TFIID and the Rpb4 and Rpb7 subunits of Pol II (ref. 16). The head has also been shown to interact with TFIIF through the Med11 subunit<sup>17</sup>. Determining the architecture of the Mediator head module is therefore vital to understanding the mechanism by which Mediator controls gene expression.

We engineered the head module to obtain crystals of sufficient quality for structure determination (Supplementary Information, section 1). In our engineered Mediator head, Med18 loop regions and the amino-terminal 108 residues of Med17 were deleted, without apparent effect on the integrity of the complex (Supplementary Fig. 1). The modified head module was labelled with selenomethionine (SeMet) and purified as described previously<sup>16</sup>. By overcoming two major technical obstacles (Supplementary Information, section 2), we produced SeMet crystals that diffract to 4.3 Å (Supplementary Table 1). The electron density map was calculated to a resolution of 4.3 Å (Supplementary Fig. 2) by SeMet single anomalous dispersion (SAD) after initial phases had been obtained using Ta<sub>6</sub>Br<sub>14</sub> and K<sub>3</sub>Ir(NO<sub>3</sub>)<sub>6</sub> derivatives (Supplementary Information, section 3).

We began identification of the individual polypeptide constituents of the Mediator head module by docking the Med18–Med20–Med8 C-terminal helix (CTH) complex structure<sup>18</sup> (Protein Data Bank ID, 2HZS) into the electron density map and then performing rigid-body refinement. The polypeptide chains of the other subunits were identified on the basis of the SeMet positions and their juxtaposition with large amino-acid side chains within ordered regions of secondary structure (Methods). This approach permitted the unambiguous assignment of all discernible elements of secondary structure in the density map to individual head module subunits (Fig. 1 and Supplementary Figs 3–5).

Our crystal structure is consistent with the molecular envelope of the head module derived at a resolution of 30–35 Å by single-particle electron microscopy analysis (Supplementary Figs 6 and 7). The head can be described in terms of three major domains, a ‘fixed jaw’, a ‘movable jaw’ and a ‘neck’ (Fig. 1 and Supplementary Figs 4 and 5), with a ‘central joint’ connecting these domains. Our X-ray structure of the head module reveals the overall architecture of the module and the domain boundaries. The domains are connected through flexible loops and linkers at the central joint.

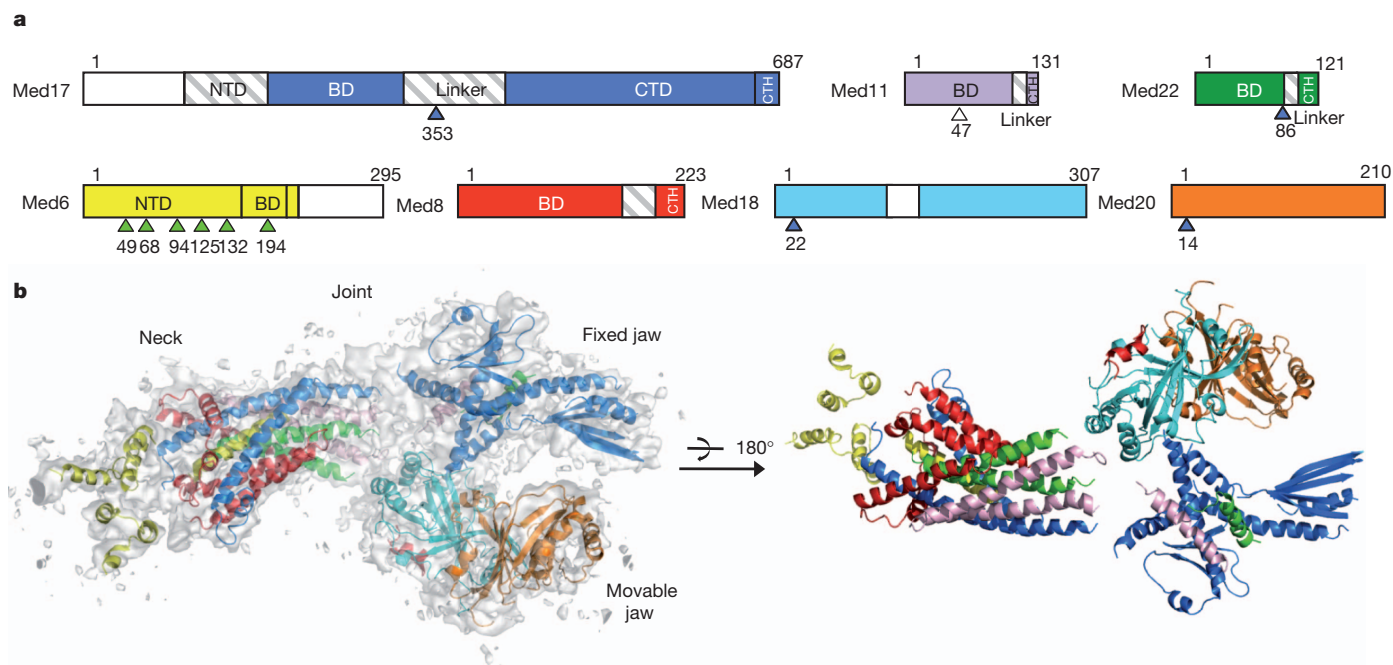
Our previous work on expression and purification of the head module suggested that Med17 has a central role in head assembly<sup>10</sup>. The work we report here extends those results through a comprehensive biochemical analysis in combination with electron microscopy, to determine the Med17 domain structure and elucidate its interactions with other head components (Supplementary Information, section 4). The results support our model of the architecture of the head module.

Assembly of the head module starts with formation of the ‘mini-head’ (Med17–Med11–Med22). Subsequently, Med8 and Med6 are added, followed by Med20–Med18 (ref. 10). Our structure shows that a four-helix bundle, built by  $\alpha$ -helices from Med11 (BH1 and BH2) and Med22 (BH1 and BH2) interact with BH2 of Med17 to form the larger helical bundle (Figs 2 and 3 and Supplementary Fig. 4). This is consistent with the observation that omission of either Med11 or Med22 leads to disassembly of the head<sup>10</sup>. Med6 interacts with the mini-head through its BH1, and Med8 serves to stabilize the central  $\alpha$ -helical bundle by surrounding the central helices. Finally, the Med18–Med20 heterodimer binds to the core-head, which is composed of five subunits (Med6, Med8, Med11, Med22 and Med17), primarily through the CTH of Med8 (Fig. 2).

The fixed jaw domain comprises the CTHs of Med11 and Med22 and the CTD of Med17. The Med11 and Med22 CTHs interact with the helical regions of the Med17 CTD. Med17 (residues 610–660) forms a  $\beta$ -sheet structure that lines the inner surface of the fixed jaw and faces the movable jaw (Fig. 3a). The Med17 CTD interacts with the loop region of Med18. The functional importance of the Med17 CTD correlates with the biochemical activity of the Head module *in vitro*, as well as phenotypic analysis *in vivo*, as loss of the Med17 CTD abolishes the transcription activity of the head module (Supplementary Fig. 12),

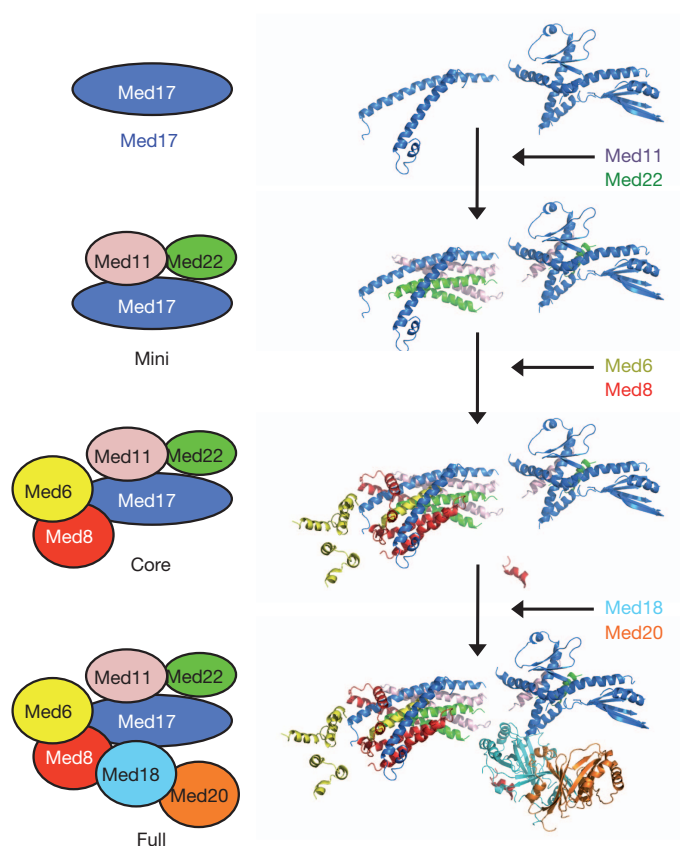
<sup>1</sup>Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 635 Barnhill Drive, Indianapolis, Indiana 46202, USA. <sup>2</sup>Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94350, USA. <sup>3</sup>Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. <sup>4</sup>Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. <sup>5</sup>European Molecular Biology Laboratory, Grenoble Outstation, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France. <sup>†</sup>Present addresses: Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15260, USA (G. Calero); School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230027, China (G. Cai).



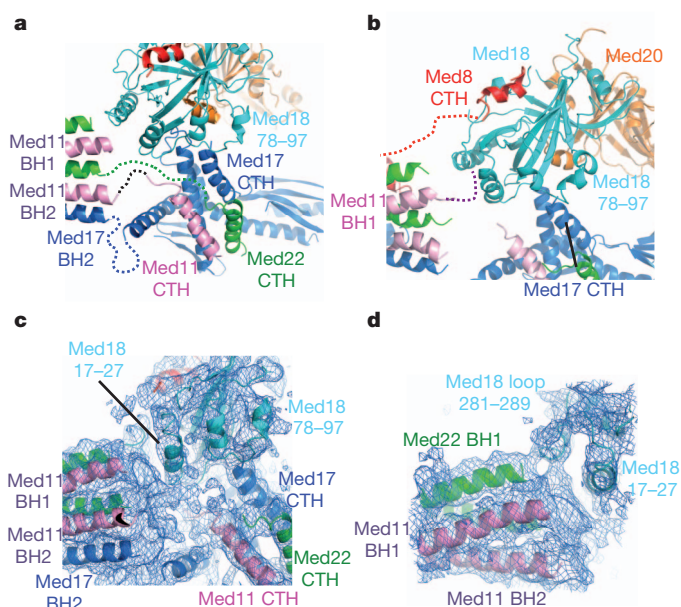


**Figure 1 | Overall structure of the Mediator head module.** **a**, Head module subunit domains. Med17 is shown in blue, Med11 in purple, Med22 in dark green, Med6 in yellow, Med8 in red, Med18 in cyan and Med20 in orange. The regions not modelled are hatched in grey and the regions not present in the crystal are shown in white. Positions of *med6<sup>ts</sup>* mutations are marked by green

arrows, *srb* suppressor mutations by blue arrows and Med11 residue 47 (Thr) by a white arrow. BD, bundle domain; CTH, C-terminal domain; NTD, N-terminal domain. **b**, A ribbon model of the Mediator head module superimposed on the experimental electron density map contoured at  $1.5\sigma$ .



**Figure 2 | Mechanism of Mediator head module complex assembly.** Models of the mini-head (Med17, Med11 and Med22) and core-head (mini-head with Med6 and Med8) modules as derived from our crystal structure of the full head module (core-head with Med18 and Med20). Diagrams of head module components (left) and corresponding structures (right) are shown.



**Figure 3 | Structures of fixed and movable jaw domains.** **a**, Fixed jaw domain interactions. The linker regions of Med17 (residues 320–420), Med11 (94–110) and Med22 (87–100) are drawn as dotted lines. **b**, Movable jaw domain interactions. Linker regions of Med8 and Med11 are drawn as dotted lines. **c**, Electron density map at the central joint region, showing density corresponding to the linker regions of Med11, Med22 and Med17. **d**, Electron density map at the junction of the Med11, Med22 and Med18 subunits. The models of Med22 BH1, Med11 BH1 and BH2, and the Med18 loop region (residues 17–27 and 281–289) are superimposed.

and all Med17 CTD deletion mutants as well as internal deletion mutants result in lethality (Supplementary Fig. 13).

The movable jaw, so called because previous electron microscopy studies<sup>16</sup> demonstrated multiple orientations of this domain with respect to the rest of the head module, is formed by the Med18–Med20–Med8 CTH complex. As for the interaction with the Med8 CTH<sup>18</sup>, our complete head module structure has revealed additional interactions with the fixed jaw and neck domains. First, the Med18 loop region formed by residues 78–97 interacts with the Med17 CTH of the fixed jaw domain (Fig. 3a, b). Second, the electron density corresponding to the N terminus of the Med11 subunit (residues 1–20) indicates an interaction with Med18 (residues 17–27 and 281–289; Fig. 3c, d). The assignment of Med11 residues 1–20 was complicated by the substitution of Ser 17 for Met 17 (Methods), and an unambiguous sequence marker is therefore lacking. However, our biochemical data (Supplementary Information, section 5) support our architectural model, in which a stable association between Med18–Med20 and the head module requires binding to Med8 and at least one additional interaction (with Med11 or Med17). The interactions with the CTH of Med17 and the NTD of Med11 are likely to be critical for the functional positioning and flexibility of the movable jaw<sup>16</sup> (Fig. 3b–d and Supplementary Fig. 6).

The neck domain has an unusual structure: a total of ten helices from five different subunits associate through the formation of a large helical bundle. The NTD of Med6 is located adjacent to the large helical bundle and consists of four  $\alpha$ -helices (Figs 1b and 4a). The helical bundle of the neck domain can be divided into two parts, a

short bundle composed of four short  $\alpha$ -helices and a long bundle composed of six long  $\alpha$ -helices. Three helices of the Med8 subunit (BH3, BH4 and BH5) seem to stabilize the assembly of both short and long bundles, and, thus, the entire neck domain structure. TATA-binding protein was reported to bind to the N-terminal 138 residues of Med8 (ref. 18), which corresponds to helices BH1 to BH5, all of which are located on the surface of the neck domain.

The organization of the helical bundle in the neck domain may produce a relatively rigid structure that could mechanically convey regulatory signals. Several observations suggest that Med6 may function as an interface between the Mediator head and middle modules, and transduce a mechanical signal from the tail or middle to the head and onto Pol II (Supplementary Information, section 6).

Mediator stimulates the phosphorylation of Ser 5 in the Pol II CTD by TFIIF (ref. 19), which promotes dissociation of Mediator from Pol II (refs 20, 21), an important step in the transition from initiation to elongation of transcription<sup>22</sup>. Our structural and biochemical data, along with relevant previous observations<sup>12,17,23,24</sup>, suggest an interaction of the Pol II CTD, the Mediator head module and TFIIF. First, mutation of Thr 47 to Ala in Med11 affects the interaction of TFIIF with the head module *in vivo*, resulting in a reduction of Pol II CTD Ser 5 phosphorylation<sup>17</sup>. Thr 47 of Med11 is located near the centre of the two symmetrical, long helical bundles of the neck, which thus could constitute the docking surface for TFIIF (Fig. 4b and Supplementary Fig. 16c). Second, three of four suppressor mutations of Pol II CTD truncation—Med17 (Gly 353 to Cys), Med22 (Asn 86 to Lys) and Med18 (Thr 22 to Ile)—map to the central joint region<sup>12,24</sup> (Supplementary Fig. 16a), suggesting that there is a functional interaction between the CTD and this portion of the head, consistent with previous observations<sup>23</sup>. Third, the head module within the Mediator/Pol II structure (Fig. 4b and Supplementary Fig. 17) is located near the base of the CTD. Finally, our biochemical data show that the head module stimulates phosphorylation of the Pol II CTD by TFIIF (Supplementary Information, section 7, and Supplementary Fig. 18). Therefore, we suggest that the head module may function as a scaffold that juxtaposes TFIIF and the Pol II CTD, thereby facilitating CTD phosphorylation (Fig. 4b). Our Mediator head structure reveals intricate interaction networks, notably the striking multi-helical bundle in the neck domain, engaging five Mediator subunits in a single structure unit. Such interactions could not have been determined from structures of individual subunits alone, nor from analysing pairwise small domain–domain interactions, but only by study of the multi-protein complex in its entirety.

## METHODS SUMMARY

**Structure determination.** Modified head module was expressed with the MultiBac system<sup>25</sup> in insect cells and purified by nickel affinity chromatography. Crystals were obtained by the hanging-drop vapour diffusion method. The structure was determined by SeMet SAD after a sufficient number (98) of SeMet sites had been identified from a combination of initial phases obtained using Ta<sub>6</sub>Br<sub>14</sub> and iridium derivatives and partial-model SAD phases.

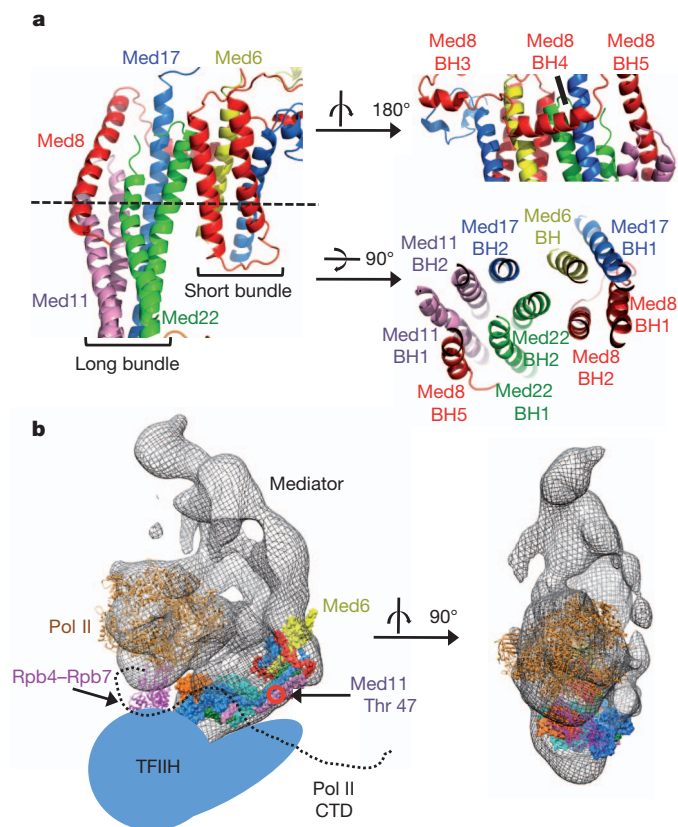
**Biochemical and electron microscopy analysis.** The Mediator head and its mutants were expressed in insect cells and purified by nickel affinity chromatography. The electron microscope images of the head module and the mutants were collected and class averages were calculated.

**In vitro assays and yeast genetics.** The *in vitro* transcription assay to assess activity of the recombinant head module and its mutant form using *srb4<sup>ts</sup>* mutant crude extract, the assay for phosphorylation of the CTD of Pol II by TFIIF, and the yeast phenotypic analysis were all done as described previously<sup>15</sup>.

Full Methods and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 24 January; accepted 28 April 2011.

Published online 3 July 2011.



**Figure 4 | Structure of the neck domain, and model of the Pol II–Mediator–TFIIF complex.** **a**, The neck domain is depicted in front (left), back (top right) and top (bottom right) views. **b**, Model of the Pol II–Mediator–TFIIF complex. Pol II and the head structures were docked into the electron microscopy map of Mediator–Pol II, shown as a mesh<sup>8</sup>. The head module is coloured as in Fig. 1b. Core Pol II is in brown, Rpb4–Rpb7 is in purple, the Pol II CTD is drawn as a black dotted line and TFIIF is shown schematically (light blue). The location of Med11 residue 47 (Thr) is indicated.

2. Conaway, R. C., Sato, S., Tomomori-Sato, C., Yao, T. & Conaway, J. W. The mammalian Mediator complex and its role in transcriptional regulation. *Trends Biochem. Sci.* **30**, 250–255 (2005).
3. Boube, M., Joulia, L., Cribbs, D. L. & Bourbon, H. M. Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. *Cell* **110**, 143–151 (2002).
4. Malik, S. & Roeder, R. G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature Rev. Genet.* **11**, 761–772 (2010).
5. Bjorklund, S. & Gustafsson, C. M. The yeast Mediator complex and its regulation. *Trends Biochem. Sci.* **30**, 240–244 (2005).
6. Guglielmi, B. *et al.* A high resolution protein interaction map of the yeast Mediator complex. *Nucleic Acids Res.* **32**, 5379–5391 (2004).
7. Asturias, F. J., Jiang, Y. W., Myers, L. C., Gustafsson, C. M. & Kornberg, R. D. Conserved structures of mediator and RNA polymerase II holoenzyme. *Science* **283**, 985–987 (1999).
8. Davis, J. A., Takagi, Y., Kornberg, R. D. & Asturias, F. A. Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol. Cell* **10**, 409–415 (2002).
9. Cai, G., Imasaki, T., Takagi, Y. & Asturias, F. J. Mediator structural conservation and implications for the regulation mechanism. *Structure* **17**, 559–567 (2009).
10. Takagi, Y. *et al.* Head module control of mediator interactions. *Mol. Cell* **23**, 355–364 (2006).
11. Nonet, M. L. & Young, R. A. Intragenic and extragenic suppressors of mutations in the heptapeptide repeat domain of *Saccharomyces cerevisiae* RNA polymerase II. *Genetics* **123**, 715–724 (1989).
12. Thompson, C. M., Koleske, A. J., Chao, D. M. & Young, R. A. A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* **73**, 1361–1375 (1993).
13. Thompson, C. M. & Young, R. A. General requirement for RNA polymerase II holoenzymes *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 4587–4590 (1995).
14. Holstege, F. C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
15. Takagi, Y. & Kornberg, R. D. Mediator as a general transcription factor. *J. Biol. Chem.* **281**, 80–89 (2006).
16. Cai, G. *et al.* Mediator head module structure and functional interactions. *Nature Struct. Mol. Biol.* **17**, 273–279 (2010).
17. Esnault, C. *et al.* Mediator-dependent recruitment of TFIH modules in preinitiation complex. *Mol. Cell* **31**, 337–346 (2008).
18. Larivière, L. *et al.* Structure and TBP binding of the Mediator head subcomplex Med8–Med18–Med20. *Nature Struct. Mol. Biol.* **13**, 895–901 (2006).
19. Kim, Y. J., Bjorklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599–608 (1994).
20. Svejstrup, J. Q. *et al.* Evidence for a mediator cycle at the initiation of transcription. *Proc. Natl Acad. Sci. USA* **94**, 6075–6078 (1997).
21. Max, T., Sogaard, M. & Svejstrup, J. Q. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *J. Biol. Chem.* **282**, 14113–14120 (2007).
22. Payne, J. M., Laybourn, P. J. & Dahmus, M. E. The transition of RNA polymerase II from initiation to elongation is associated with phosphorylation of the carboxyl-terminal domain of subunit IIa. *J. Biol. Chem.* **264**, 19621–19629 (1989).
23. Kang, J. S. *et al.* The structural and functional organization of the yeast mediator complex. *J. Biol. Chem.* **276**, 42003–42010 (2001).
24. Koleske, A. J., Buratowski, S., Nonet, M. & Young, R. A. A novel transcription factor reveals a functional link between the RNA polymerase II CTD and TFIID. *Cell* **69**, 883–894 (1992).
25. Fitzgerald, D. J. *et al.* Protein complex expression by using multigene baculoviral vectors. *Nature Methods* **3**, 1021–1032 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank L. Messerle for providing the Ta<sub>6</sub>Br<sub>14</sub> metal cluster, T. Hurley for reading the manuscript, M. Georgiadis for discussions, C. Kaplan for his advice on yeast genetics, T. Earnest for giving us beam time and L. Fabrizio for assisting with the N-terminal sequence analysis. We thank the CCP4 summer school, funded by the NCI (Y1-CO-1020), and NIGMS (Y1-GM-1104) for their assistance with the twinning data analysis. Y.T. thanks the instructors on 'The X-Ray Methods Course' at Cold Spring Harbor Laboratory. This work was supported by US National Science Foundation grant MCB 0843026 (Y.T.); the American Heart Association 0735395N (Y.T.); a Human Frontier Science Program long-term fellowship (T.I.); NIH grants R01 GM67167 (F.J.A.) and GM36659 (R.D.K.); NCI Cancer Center Support Grant P30 CA08748 (to the MSKCC Microchemistry and Proteomics Core Laboratory); and the European Commission Framework Program 7 projects INSTRUCT and P-CUBE (I.B.). X-ray data were collected at the GM/CA-CAT at the Advanced Photon Source, Argonne National Laboratory. GM/CA-CAT is funded by the NIH (Y1-CO-1020 and Y1-GM-1104) and the Advanced Photon Source is supported by the DOE (DE-AC02-06CH11357). Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, supported by the DOE and the NIH, and at the Advanced Light Source, supported by the DOE (DE-AC02-05CH11231).

**Author Contributions** T.I., I.B. and Y.T. implemented the MultiBac system. T.I. was mainly responsible for protein complex preparation, crystallization, data collection, data analysis and model building in collaboration with Y.T. T.I., H.E.-B. and P.T. carried out mass spectroscopy analysis. G. Calero, G.L.K. and Y.T. carried out the initial crystallization and data collection, supervised by R.D.K., Y.T., T.I. and F.C. designed and carried out expression of the mutant head modules and their biochemical characterization. Y.T. and K.Y. designed and carried out the yeast genetic experiment. Y.T. carried out the *in vitro* transcription assay and the CTD kinase assay; G. Cai, K.-L.T. and F.J.A. carried out the electron microscopy study on the head module and its mutants. T.I., F.J.A. and Y.T. discussed and interpreted all results. Y.T. supervised the X-ray, biochemical and yeast genetic work, and wrote the manuscript in collaboration with T.I., I.B., F.J.A. and R.D.K.

**Author Information** Coordinates and structure factors have been deposited in the Protein Data Bank under accession code 3RJ1. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to Y.T. ([ytakagi@iupui.edu](mailto:ytakagi@iupui.edu)).



## METHODS

**Construction of vectors.** All the vectors used in this study are summarized in Supplementary Tables 3 and 4. For expression of the modified head module for crystallization, DNA sequences corresponding to residues 1–108 of Med17 were removed from vector pFL-10xHis-Med17 (ref. 16; pYT49) by the SLIC method<sup>26</sup>, yielding pFL-10xHis-Med17 (109–687) (pYT171). DNA sequences corresponding to residues 109–140 and 71–156 of Med18, respectively, were removed from vector pSPL-Med18-Med20 (ref. 16; pYT75), yielding pSPL-Med20-Med18 ( $\Delta$ 109–140) (pYT115) and pSPL-Med20-Med18 ( $\Delta$ 71–156) (pYT114). To eliminate an alternative translation start site, the Met (residue 17) of Med11 was mutated to Ser (pYT311). Finally, the transfer vector for the modified head module was generated by fusing three vectors, pYT171, pYT114 and pUCDM-Med6-Med22-Med11-Med8 (pYT120) by Cre/LoxP recombination as previously described<sup>25</sup>.

DNA sequences corresponding to residues 1–16 of Med11 were removed from the vector pYT111 by SLIC, yielding the vector pUCDM-Med22-Med11 ( $\Delta$ 1–16) (pYT147). Fusion of pYT171, pYT147 and pYT120 with either pYT114 or, alternatively, pYT115 generated the expression vectors for a series of double Med18–Med11 partial head module deletion mutants.

The constructs for Med17 mutagenesis were generated as follows. BamHI and HindIII fragments corresponding to the C-terminal deletion mutants of Med17 were generated by first introducing a stop codon and a HindIII site (TAAAGCTT) into pBacPAK9-10His-SRB4 (MED17) vector<sup>10</sup> adjacent to the sequences corresponding to residues 108, 200, 300, 400, 500 and 600 of Med17, by using the QuickChange mutagenesis kit (Stratagene), followed by BamHI and HindIII digestion and gel purification. The respective purified fragments were cloned into the BamHI and HindIII sites of pFL vector<sup>25</sup>, yielding vectors pYT165 to pYT170 (Supplementary Table 3). The N-terminal deletion, as well as the internal deletion mutant constructs of Med17, were generated by removing DNA sequences corresponding to residues 1–108, 1–201, 1–302, 1–400, 101–200, 201–300 and 301–400 from pFL-10xHis-Med17 by the SLIC method, yielding the respective vectors pYT183 to pYT186 and pYT289 to pYT291 (Supplementary Table 3). These vectors were fused with pUCDM-Med6-Med22-Med18-Med20-Med11-Med8 (pYT151), yielding vectors encoding for head modules comprising Med17 mutant forms. The vector pYT151 was created by two rounds of sequential cloning of PmeI and the AvrII fragments containing Med18–Med20 and Med22–Med11 into SpeI and NruI sites of pUCDM-Med6-Med8 (pYT110).

Introduction of the deletion mutations into the yeast shuttle vector pCT127, carrying the wild-type *MED17* gene, was also carried out by SLIC. The yeast shuttle vectors used in this study are listed in Supplementary Table 4.

**Expression and purification of the head module and mutants.** Expression and purification of the recombinant head module, the mutant forms and the subcomplex was carried out in insect cells using the MultiBac system<sup>25</sup>. Production of high-titer viruses in Sf9 cells, and expression and purification of recombinant head modules of Mediator and its mutant forms was carried out as described previously<sup>16</sup>.

Preparation of SeMet labelled Head module will be described elsewhere (T.I. *et al.*, manuscript in preparation). Briefly, the insect cells were cultured in Met-free medium (Expression Systems) overnight before baculovirus infection. L-selenomethionine (20 mg l<sup>-1</sup>; Sigma-Aldrich) was added at sequential 24-h intervals. Cells were collected 96 h after infection. SeMet-labelled complex was purified as described above.

**Limited proteolysis and identification of the peptide fragments.** A total of 135  $\mu$ g of the recombinant head module was incubated at 37 °C with chymotrypsin (Sigma-Aldrich) at a final concentration of 0.01 mg ml<sup>-1</sup> in a volume of 150  $\mu$ l in buffer A (20 mM Tris-HCl (pH 8.0), 100 mM NaCl and 1 mM DTT). Aliquots (20  $\mu$ l) were taken at 0, 5, 10, 30 and 60 min, and 15  $\mu$ l of PMSF stock solution (100 mg ml<sup>-1</sup>) was added to stop the reaction by inhibiting the protease. Aliquots were applied to 12.5% SDS–PAGE and transferred onto a Sequi-Blot PVDF membrane (Bio-Rad). Protein bands were stained by Coomassie blue (R-250). Protein bands resulting from proteolysis during the time course were identified, excised and subjected to Edman degradation using a Procise 494 instrument from Applied Biosystems as previously described<sup>27</sup>. Stepwise-liberated PTH-amino acids were identified using an ‘on-line’ HPLC system (Applied Biosystems) equipped with a PTH C18 (2.1  $\times$  220 mm; 5- $\mu$ m particle size) column (Applied Biosystems).

**Crystallization and data collection.** Crystals were obtained at 293 K by hanging-drop vapour diffusion against a reservoir solution of 0.1 M Tris-HCl (pH 7.6) containing 10–12.5% (w/v) PEG-6K and 0.4 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. Crystals were transferred into the reservoir solution containing 25% triethylene glycol (TEG). The crystals were flash-frozen for data collection at 100 K. SDS–PAGE analysis of the dissolved crystals confirmed the presence of all seven subunits. However, *in situ* proteolysis resulted in about 10% of the Med17 subunits being shortened at the N terminus by 76 residues and almost 100% of the Med6 subunits being shortened at the C terminus by 80 residues (Supplementary Fig. 1). Diffraction data were collected at beamline 23ID

at the Advanced Photon Source (APS) at Argonne National Laboratory. All diffraction data were processed with HKL2000<sup>28</sup>. Twinning rates of the data sets were analysed using program PHENIX XTRIAGE<sup>29</sup>.

**Structure determination of the Mediator head module.** Initial phases were determined by two approaches: Ta<sub>6</sub>Br<sub>14</sub> single isomorphous replacement with anomalous scattering (SIRAS) and iridium single anomalous dispersion (SAD). Ta<sub>6</sub>Br<sub>14</sub> derivative crystals were prepared by soaking the native head module crystals in reservoir solution containing 1 mM Ta<sub>6</sub>Br<sub>14</sub>. The initial phase was determined by SIRAS at a resolution of 7.5 Å. Density modification using the program PARROT extended the phase resolution to 4.3 Å using the SeMet data set. Iridium derivatives of the crystal of Mediator head module were prepared by soaking the crystals in crystallization reservoir solution containing 10 mM K<sub>3</sub>Ir(NO<sub>3</sub>)<sub>6</sub>. The initial iridium phase was obtained by SAD using the programs SHELXD and PHASER<sup>30,31</sup>. The phase was extended followed by density modification by program PARROT<sup>32</sup> with the SeMet data set. However, the maps obtained at this stage were not yet interpretable.

To improve the maps, we used them together and applied the following methods: (i) location of SeMet sites in the crystal; (ii) non-crystallographic symmetry (NCS), averaging between three molecules in the NCS using the program DM<sup>33</sup>; (iii) partial model building into the clearly discernible rod-like electron density from  $\alpha$ -helices, followed by rigid-body refinement using the programs COOT and REFMAC5<sup>34</sup>; and (iv) re-calculating phases by SeMet SAD phasing with PHASER, using the partial model and SeMet positions. Iterative rounds combining these procedures were performed until the model covered all interpretable secondary structure elements. Eventually, we could identify 98 SeMet sites. To minimize model bias, phases were re-calculated by SeMet SAD with PHASER, using only the positions of these 98 selenium sites, and these improved SAD phases guided the final model-building steps.

**Model building and refinement.** Assignment of polypeptide identities was carried out as follows. The published structure<sup>18</sup> of Med18–Med20–Med8 (CTH) (PDB ID, 2HZS) was manually docked into the electron density map, followed by rigid-body refinement by using COOT. Then the  $\alpha$ -helices for the further polypeptides of the head module were manually built, and connected. Next,  $\beta$ -sheets were manually built into the unassigned structured regions in the electron density, which corresponded to the neck and fixed jaw domains. Subsequently, we began assigning the polypeptide identities at the neck domain. We tracked specific SeMet labelling patterns dictated by the presence of Met in the primary sequences of the polypeptides, and the presence of bulky regions corresponding to aromatic residue positions, as markers. We used secondary structure predictions for additional guidance. First, Med8 BH1, Med17 BH1 and Med22 BH2 were identified in  $\alpha$ -helical bundle regions in the neck domain from their primary-sequence-specific, unique SeMet labelling pattern: these regions all contain more than two SeMet peaks and the spacing of SeMet peaks was consistent with the corresponding amino-acid sequences in the subunits. This assignment was consistent with the secondary structure predictions indicating  $\alpha$ -helical structure. The remaining Med8 residues (60–170), as well as Med22 BH1, were assigned by tracing from the Med8 BH1 helix back to the Med8 C terminus, and by tracing from the Med22 BH2 helix back to the Med22 N terminus. This assignment was validated by the fact that their Met locations aligned with anomalous peaks on the experimental map. Next, we identified the NTD of Med6 based on its unique SeMet positions, and also identified Med6 BH based on a specific location of SeMet (Met48), a bulky aromatic ring (Phe52) (Supplementary Fig. 19a) and continuity from the NTD of Med6, consistent with secondary structure predictions. The C-terminal 80 residues of Med6 were proteolyzed in the crystals (Supplementary Fig. 1). Consequently, no density was found corresponding to the C terminus of Med6. We traced Med17 BH1, and identified the longest helix in the neck domain as the BH2 helix of Med17 on the basis of a single SeMet (Met313) and the aromatic side chain of Tyr269 (Supplementary Fig. 19b); this assignment was also consistent with secondary structure predictions. Finally, the two remaining continuous  $\alpha$ -helices in the neck domain were identified as Med11 BH1 and Med11 BH2 because of one unique SeMet position of Med11. This assignment matches perfectly to the secondary structure prediction as well.

Next we focused on the fixed jaw domain. By subtracting the polypeptides already assigned to the neck and the movable jaw (see above), the fixed jaw should only contain the C-terminal regions of subunits Med11, Med17 and Med22. First, on the basis of continuity, SeMet position (Met422), aromatic ring position (Tyr423) (Supplementary Fig. 19c), and secondary structure prediction, we identified helix 420–455,  $\beta$ -sheet 456–480 and helices 496–523 and 540–570 of Med17. The remainder of the electron density in this region was continuous, and thus enabled us to trace Med17 completely to its C terminus. We identified the Med17 CTH and  $\beta$ -sheet with  $\alpha$ -helix 600–608 on the basis of the SeMet positions and  $\alpha$ -helix length from secondary structure prediction. Finally, we assigned two remaining helices: Med11 CTH was identified from the presence of one SeMet peak, and we assigned the last helix to Med22 CTH, which entirely lacks SeMet.

Initially, all models were refined using the program CNS DEN<sup>35</sup>, refinement with strong NCS restraints between the three independent complexes in the asymmetric unit, and twinning refinement. Then the model was refined using PHENIX with NCS restraints and a single refined group isotropic temperature factor for each subunit, Ramachandran restraint, TLS refinement and twinning refinement. The geometry of the final model is good, with 91.3%, 8.0%, 0.7% of the amino-acid residues in the most favoured, allowed, and disallowed regions of the Ramachandran plot, respectively. All structural illustrations and electron density maps were prepared with PYMOL (<http://www.pymol.org/>) and COOT. PSIPRED was used for secondary structure prediction<sup>36</sup>.

**Docking of the X-ray structure into the electron microscopy map.** The model of 12-subunit Pol II was docked into the Mediator–Pol II holoenzyme structure<sup>8</sup> followed by docking of the X-ray model of the head module into the density corresponding to the Mediator head module, using the program CHIMERA<sup>37</sup>.

**Electron microscopy sample preparation, data collection and image analysis.** We diluted purified head module deletion mutants in buffer containing 25 mM KCl, 25 mM Tris-HCl (pH 7.8) and 5 mM DTT. For preparation of all electron microscopy samples, about 3 µl of protein solution was applied to a carbon-coated Maxtaform, 300-mesh Cu/Rh EM specimen grid (Ted Pella) freshly glow-discharged in the presence of amylamine. The particles were then preserved by staining with a 2.0% (w/w) uranyl acetate solution using the sandwich carbon layer technique<sup>38,39</sup>. The images were recorded under low-dose conditions using a Tecnai Spirit (Philips/FEI) microscope equipped with a LaB6 filament and operating at an accelerating voltage of 120 kV. Images were recorded on a Tietz (TVIPS) CCD camera at ×42,000 magnification and approximately 1-µm under-focus, resulting in a final pixel size corresponding to 5.06 Å.

The images were initially analysed using the ml\_align2d program, a multi-reference, two-dimensional alignment routine with a maximum-likelihood target function<sup>40</sup> implemented in the XMIPP package<sup>41</sup>. Averages derived from the ml\_align2d program were used to run iterative alternating rounds of supervised multi-reference alignment/classification and reference-free alignment as described previously<sup>42</sup> to improve the homogeneity of the image classes.

**In vitro transcription and the CTD phosphorylation assays.** The *in vitro* transcription assay to assess activity of the recombinant head module and its mutant form using *srb4*<sup>45</sup> mutant crude extract was performed as described previously<sup>10</sup>. Quantification of transcripts on an absolute scale was performed using a FLA-5100 FUJIFILM fluorescent image analyser and the MultiGauge software package after

addition of 1 nCi of <sup>32</sup>P UTP to the gel 5 min before the end of the run. The CTD phosphorylation assay was performed as previously described<sup>15</sup>.

**Yeast phenotypic analysis.** The shuttle vectors carrying the MED17 mutations are described in Supplementary Table 4. The shuttle vectors were introduced into yeast strain Z572 by plasmid shuffling, and grown on SC medium containing 5-FOA at 30 °C as previously described<sup>15</sup>.

26. Li, M. Z. & Elledge, S. J. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. *Nature Methods* **4**, 251–256 (2007).
27. Tempst, P., Geromanos, S., Elicone, C. & Erdjument-Bromage, H. Improvements in microsequencer performance for low picomole sequence analysis. *Methods* **6**, 248–261 (1994).
28. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
29. Adams, P. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
30. Sheldrick, G. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
31. McCoy, A. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
32. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D* **66**, 470–478 (2010).
33. Cowtan, K. Modified phased translation functions and their application to molecular-fragment location. *Acta Crystallogr. D* **54**, 750–756 (1998).
34. Emsley, P., Lohkamp, B., Scott, W. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
35. Schröder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464**, 1218–1222 (2010).
36. Bryson, K. *et al.* Protein structure prediction servers at University College London. *Nucleic Acids Res.* **33**, W36–W38 (2005).
37. Pettersen, E. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
38. Stoffer, G. & Stoffer-Meilicke, M. *The Ultrastructure of Macromolecular Complexes Studied with Antibodies* 409–455 (De Gruyter, 1983).
39. Tischendorf, G. W., Zeichhardt, H. & Stoffer, G. Determination of the location of proteins L14, L17, L18, L19, L22, L23 on the surface of the 50S ribosomal subunit of *Escherichia coli* by immune electron microscopy. *Mol. Gen. Genet.* **134**, 187–208 (1974).
40. Scheres, S. H. *et al.* Maximum-likelihood multi-reference refinement for electron microscopy images. *J. Mol. Biol.* **348**, 139–149 (2005).
41. Sorzano, C. O. *et al.* XMIPP: a new generation of an open-source image processing package for electron microscopy. *J. Struct. Biol.* **148**, 194–204 (2004).
42. Brignole, E. J., Smith, S. & Asturias, F. J. Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nature Struct. Mol. Biol.* **16**, 190–197 (2009).

# Coordination of DNA replication and histone modification by the Rik1–Dos2 complex

Fei Li<sup>1</sup>, Rob Martienssen<sup>2</sup> & W. Zacheus Cande<sup>3</sup>

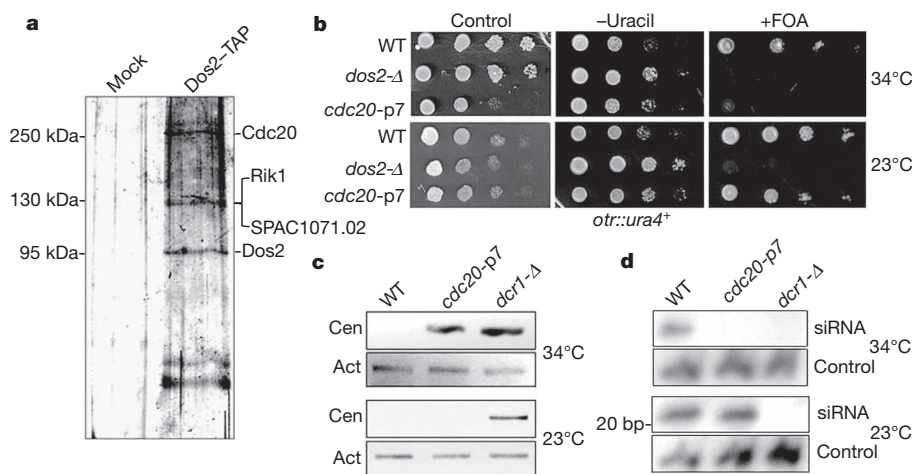
Histone modification marks have an important role in many chromatin processes<sup>1,2</sup>. During DNA replication, both heterochromatin and euchromatin are disrupted ahead of the replication fork and are then reassembled into their original epigenetic states behind the fork<sup>3,4</sup>. How histone marks are accurately inherited from generation to generation is still poorly understood. In fission yeast (*Schizosaccharomyces pombe*), RNA interference (RNAi)-mediated histone methylation is cell cycle regulated. Centromeric repeats are transiently transcribed in the S phase of the cell cycle and are processed into short interfering RNAs (siRNAs) by the complexes RITS (RNA-induced initiation of transcriptional gene silencing) and RDRC (RNA-directed RNA polymerase complex)<sup>5–7</sup>. The small RNAs together with silencing factors—including Dos1 (also known as Clr8 and Raf1), Dos2 (also known as Clr7 and Raf2), Rik1 and Lid2—promote heterochromatic methylation of histone H3 at lysine 9 (H3K9) by a histone methyltransferase, Clr4 (refs 8–13). The methylation of H3K9 provides a binding site for Swi6, a structural and functional homologue of metazoan heterochromatin protein 1 (HP1)<sup>14</sup>. Here we characterize a silencing complex in fission yeast that contains Dos2, Rik1, Mms19 and Cdc20 (the catalytic subunit of DNA polymerase- $\epsilon$ ). This complex regulates RNA polymerase II (RNA Pol II) activity in heterochromatin and is required for DNA replication and heterochromatin assembly. Our findings provide a molecular link between DNA replication and histone methylation, shedding light on how epigenetic marks are transmitted during each cell cycle.

To explore the role of Dos2 in heterochromatin assembly, we sought to identify Dos2-associated proteins by using tandem affinity

purification (TAP). Mass spectrometry analysis of Dos2 purified by the TAP method uncovered two new interacting proteins, in addition to Rik1: Cdc20 (ref. 15); and a previously uncharacterized protein, SPAC1071.02 (Fig. 1a). SPAC1071.02 is highly conserved (Supplementary Fig. 1). Its homologue in budding yeast (*Saccharomyces cerevisiae*) is MMS19 (ref. 16), so we named the fission yeast protein Mms19. The interactions of Dos2 with Mms19 and Cdc20 were confirmed by co-immunoprecipitation experiments (Supplementary Figs 2 and 3).

Cdc20 is a conserved DNA polymerase- $\epsilon$  subunit with extensive homology to its counterparts in humans and budding yeast. Cdc20 regulates the elongation of the leading strand during DNA replication, shortly after initiation, and is essential for cell viability<sup>15</sup>. To test whether Cdc20 is required for heterochromatin silencing, we used a temperature-sensitive mutant allele, *cdc20-p7*. At 37 °C, mutant cells arrest in early S phase. We crossed these mutants into the *otr::ura4<sup>+</sup>* background and performed a silencing assay. We found that, at a non-restrictive temperature, 34 °C, the mutant cells grew poorly on control medium compared with wild-type (WT) cells (Fig. 1b), probably as a result of the replication abnormality. However, on medium lacking uracil, the mutants had more robust growth than WT cells. By contrast, on 5-fluoroarotic acid (5-FOA)-containing medium, the mutant cells had little growth, demonstrating that centromeric silencing was partially compromised. These results indicate that complete silencing of heterochromatin requires Cdc20.

In WT cells, heterochromatin transcripts are quickly processed by the RNAi machinery, but in RNAi-processing defective mutants, such as *dcr1-Δ*, these transcripts are readily detectable<sup>17</sup>. In *cdc20-p7* mutant



**Figure 1 | Cdc20 is essential for transcriptional silencing and siRNA generation.** **a**, Protein extracts from a Dos2–TAP strain or an untagged control strain (mock) were purified by the TAP method. Purified products were separated by SDS–PAGE and visualized by silver staining. kDa, kilodalton. **b**, Growth assay of serial dilutions (left to right) of strains carrying *ura4<sup>+</sup>* inserted at the pericentromeric *otr* region, grown on control medium, medium

lacking uracil (–uracil) or counter-selective 5-FOA medium (+FOA) incubated at 23 °C or 34 °C. **c**, Accumulation of centromeric transcripts analysed by RT–PCR in strains incubated at the indicated temperatures. Act, *act1<sup>+</sup>* control; Cen, centromeric transcripts. **d**, Analysis of siRNAs corresponding to centromeric repeats by northern blotting. Control, snoR69 as a loading control.

<sup>1</sup>Department of Biology, New York University, New York, New York 10003, USA. <sup>2</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. <sup>3</sup>Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA.



cells incubated at 34 °C, similar to *dcrl-Δ* mutants, pericentromeric transcripts accumulated (Fig. 1c). We then examined the amount of small RNA in the *cdc20-p7* mutant by northern blotting. Because RNAi is temperature sensitive, the abundance of small RNAs in WT cells was considerably reduced at 34 °C but was still detectable<sup>6</sup> (Fig. 1d). In the *cdc20-p7* mutant, however, siRNAs were completely absent (Fig. 1d), showing that Cdc20 promotes siRNA generation.

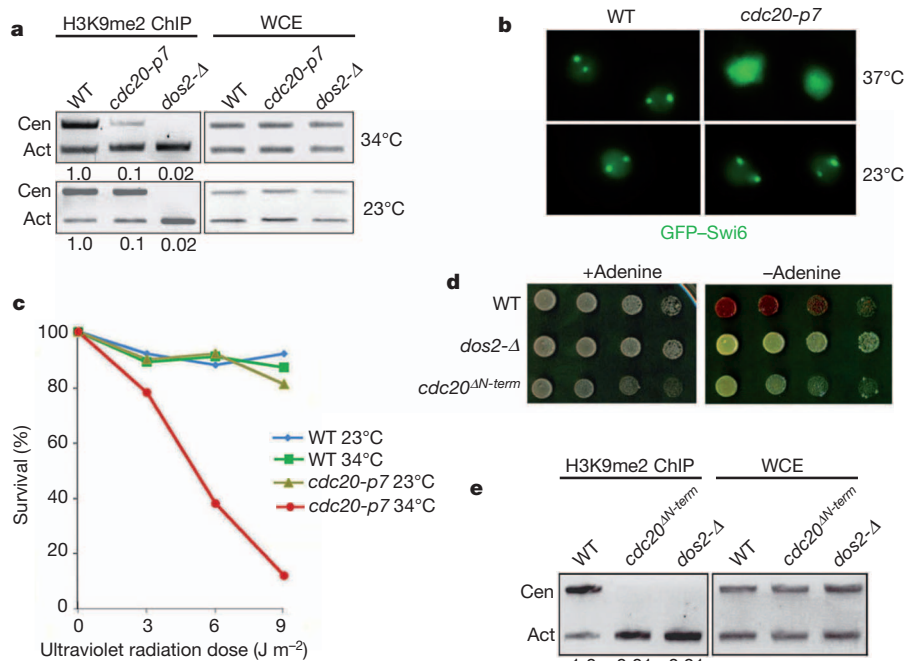
To further determine how heterochromatin structure is affected by disruption of Cdc20, we examined H3K9 methylation and Swi6 distribution in the *cdc20-p7* cells grown at 34 °C and 23 °C. Pericentromeric H3K9 methylation was significantly reduced at the higher temperature (Fig. 2a) and the association of Swi6 was also decreased (Supplementary Fig. 4), both of which are consistent with the heterochromatin defect shown by the silencing assay. We also assessed the delocalization of green fluorescent protein (GFP)–Swi6 fusion proteins to determine loss of heterochromatin, because the GFP–Swi6 pattern is unchanged in RNAi mutants<sup>18</sup>. We found that 53% of *cdc20-p7* cells at 34 °C, and more than 70% at 37 °C, had a diffuse GFP–Swi6 pattern, a defect similar to the *dos2-Δ* mutant (Supplementary Fig. 5 and Fig. 2b). WT cells incubated at the elevated temperatures did not show severe Swi6 delocalization (Fig. 2b), demonstrating that heterochromatin formation requires Cdc20. Because heterochromatin formation is mediated by both RNAi-dependent and RNAi-independent pathways<sup>19</sup>, the silencing abnormality in *cdc20-p7* mutants indicated that Cdc20 functions at an early stage of heterochromatin assembly.

Mms19, another Dos2-interacting factor, is a conserved protein that contains a HEAT repeat domain (Supplementary Fig. 1). Studies of its homologues in budding yeast and humans show that they function as regulators of the transcription factor TFIID, participating in the initiation of RNA-Pol-II-mediated transcription<sup>16,20</sup>. Interestingly, human *MMS19* is also required for chromosome segregation<sup>21</sup>. To

study the role of Mms19 in fission yeast, we first examined its distribution using a GFP-tagged version of Mms19 and found that the GFP signal was predominantly nuclear, consistent with its potential role as a transcriptional regulator (Supplementary Fig. 6). To elucidate the function of Mms19, we created an *mms19*-null (*mms19-Δ*) mutant. This mutant grew slower than the WT strain but was viable, indicating that *mms19*<sup>+</sup> is not an essential gene. Similar to the budding yeast *mms19* mutant, the growth of *mms19-Δ* required methionine (Supplementary Fig. 7).

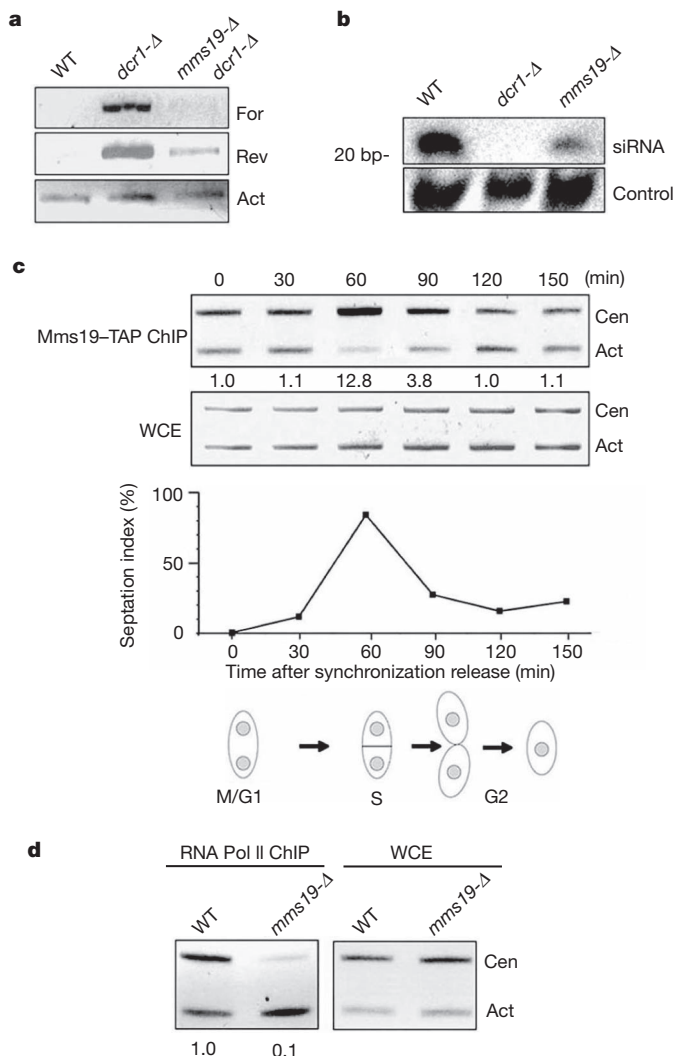
Because Mms19 homologues in other organisms associate with the TFIID complex, we speculated that Mms19 might be involved in RNA-Pol-II-mediated transcription in heterochromatic regions. To address this possibility, we directly examined centromeric transcription by using PCR with reverse transcription (RT-PCR). Centromeric transcripts were abundant in siRNA-processing mutants, such as *dcrl-Δ*, but it was difficult to detect them in the WT strain<sup>17</sup> (Supplementary Fig. 8). In contrast, these transcripts were not discernible in the *mms19-Δ* mutant by using RT-PCR, similarly to the WT strain (Supplementary Fig. 8), and their abundance was greatly reduced in a *dcrl-Δ mms19-Δ* double mutant (Fig. 3a). We reasoned that, as a result of the reduction in primary siRNA transcripts, centromeric siRNA levels might also be decreased. To test this, RNA extracted from the *mms19-Δ* mutant was probed for centromeric siRNA by northern blotting. These siRNAs were present but less abundant than in the WT strain (Fig. 3b). These data further demonstrate that Mms19 regulates centromeric transcription.

Coincident with heterochromatin expression, RNA Pol II is preferentially restricted to heterochromatin in S phase<sup>5</sup>. To further elucidate the role of Mms19 in this process, we investigated how Mms19 associates with heterochromatin during the cell cycle. After release from synchronization, cells carrying Mms19–TAP were collected at different



**Figure 2 | Heterochromatin abnormality is coupled to DNA replication defects.** **a**, Pericentromeric H3K9 methylation in *cdc20-p7* was significantly lost at 34 °C but not at 23 °C. A ChIP assay was performed with an antibody specific for dimethylated H3K9 (H3K9me2). DNA from precipitates was analysed by competitive PCR in which one set of primers amplified the centromeric *dh* repeat (Cen) and another amplified the control gene *act1*<sup>+</sup> (Act). The relative fold enrichment (indicated below each lane) was calculated by comparing the ratios of heterochromatin signals to control signals in the ChIP and whole cell extract (WCE) fractions. The value for WT cells was set to 1.0. **b**, Fluorescent images of GFP–Swi6-carrying strains incubated at 23 °C or 37 °C. GFP–Swi6 localizes to heterochromatin regions in the form of two to

four green foci in WT cells in interphase. This localization was lost in more than 70% of *cdc20-p7* cells at 37 °C. **c**, Ultraviolet radiation sensitivity of the *cdc20-p7* mutant and the WT strain at 23 °C or 34 °C. **d**, Colony colour silencing assay of strains carrying *ade6*<sup>+</sup> inserted at the pericentromeric *otr* region, grown at 23 °C on YES medium with or without further supplementation with adenine. Repression of *ade6*<sup>+</sup> expression results in a red or pink colour; when transcriptional silencing does not occur, the colonies appear white. **e**, Analysis of pericentromeric H3K9 methylation in the indicated strains. A ChIP assay was carried out with an antibody specific for H3K9me2, followed by PCR, as in **a**. The relative fold enrichment of the methylation is indicated below each lane.



**Figure 3 | Mms19 is required for RNA-Pol-II-mediated transcription of heterochromatin.** **a**, Strand-specific RT-PCR analysis of the accumulation of transcripts from centromeric *dh* repeats. For, forward strand; Rev, reverse strand. **b**, Analysis of siRNA corresponding to centromeric *dg-dh* repeats by northern blotting. bp, base pairs; control, snoR69 as a loading control. **c**, Mms19 preferentially associates with heterochromatin during S phase. Protein extracts were prepared from synchronized cells carrying Mms19-TAP, which were then analysed at various time points after synchronization release by using ChIP with an antibody specific for TAP, followed by PCR as in Fig. 2a. The relative fold enrichment of Mms19 is indicated below each lane. Cell cycle progression was monitored by a septation index, which is also illustrated diagrammatically. **d**, ChIP analysis of RNA Pol II accumulation in pericentromeric heterochromatin in cells synchronized in S phase, followed by PCR, as in c. The relative fold enrichment of RNA Pol II is indicated below each lane.

stages of the cell cycle. We found that Mms19 preferentially associated with heterochromatin in S phase (Fig. 3c), concurrently with the enrichment of RNA Pol II in heterochromatin. We then investigated how Mms19 affects the RNA Pol II distribution in heterochromatin at this stage. Using chromatin immunoprecipitation (ChIP) with an antibody specific for RNA Pol II, we found that the RNA Pol II accumulation in S phase was reduced considerably in the *mms19-Δ* mutant (Fig. 3d). Furthermore, Mms19 was found to physically associate with RNA Pol II (Supplementary Fig. 9). Together, our results suggest that Mms19 is a transcriptional activator that is required for the RNA-Pol-II-mediated transcription of heterochromatin.

To gain further insight into how Cdc20 interacts with Dos2 and Mms19, we created a haemagglutinin (HA)-tagged version of the

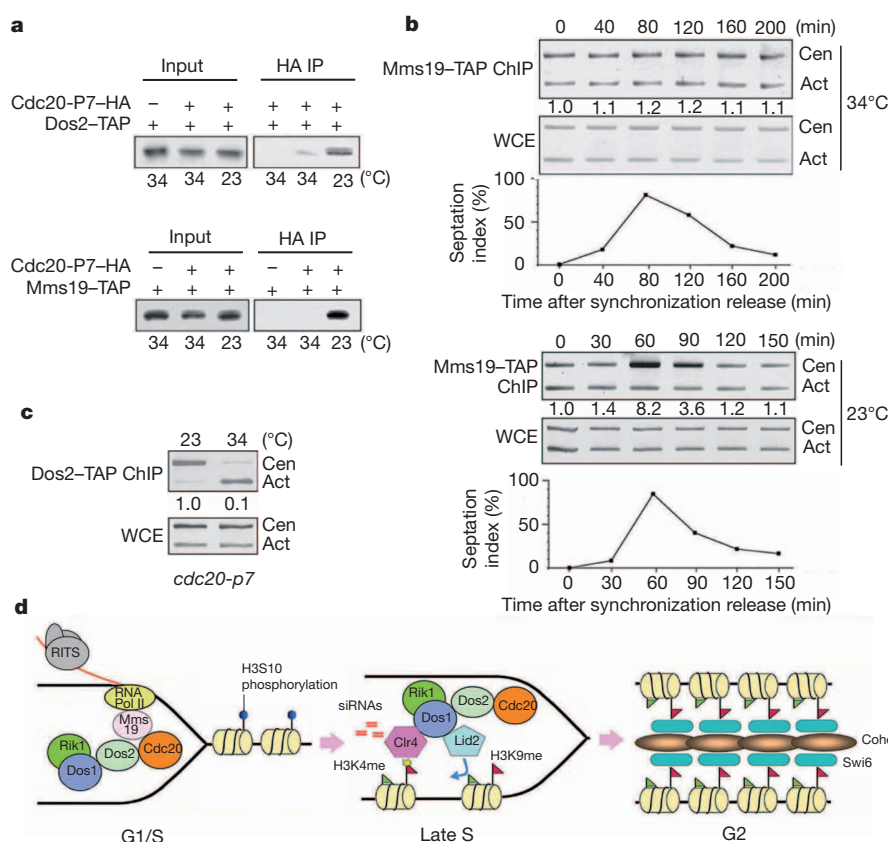
mutant gene from the *cdc20-p7* strain. Co-immunoprecipitation experiments showed that Cdc20-P7-HA maintained its association with Dos2 and Mms19 at 23 °C; however, these interactions were lost at 34 °C, indicating that the point mutation reduces the interactions at the elevated temperature (Fig. 4a). We also investigated the association between Mms19 and heterochromatin by using ChIP in synchronized *cdc20-p7* cells released from metaphase. At 23 °C, there was a clear peak in Mms19 enrichment in the mutant during S phase, which occurred ~60 min after release from synchronization; however, this accumulation of Mms19 was not observed when the temperature was elevated to 34 °C, indicating that Cdc20 is required for the association of Mms19 with heterochromatin (Fig. 4b). A previous report showed that Dos2 and Rik1 start to accumulate in heterochromatin in S phase<sup>5</sup>. We therefore assessed whether Cdc20 affects the recruitment of these two silencing factors. Using ChIP with antibodies specific for TAP or a Myc tag, we found that Dos2-TAP and Rik1-Myc are enriched in S phase at 23 °C, consistent with the previous study; however, the association is diminished at 34 °C (Fig. 4c and Supplementary Fig. 10). These results indicate that Cdc20 is required for the recruitment of Dos2 and Rik1 to heterochromatin. Interestingly, heterochromatin silencing is also partly compromised in strains with mutations in two different DNA polymerase- $\alpha$  subunits<sup>22,23</sup>. As DNA polymerase- $\alpha$  primase is required before elongation by DNA polymerase- $\epsilon$  (of which Cdc20 is a subunit), it is possible that the interaction of the Rik1-containing complex with Cdc20 underlies this silencing defect.

We reasoned that the loss of silencing in the mutant may be linked to the impaired DNA replication. Indeed, *cdc20-p7* cells grew much more slowly at 34 °C than at 23 °C and had an extended S phase (Fig. 1b). The replication state of a strain can be tested by assessing the efficiency of replication recovery after ultraviolet-radiation-induced damage. We found that the *cdc20-p7* mutant was highly sensitive to ultraviolet radiation at 34 °C but not at 23 °C (Fig. 2c). Furthermore, heterochromatin fragments that contain ARS elements could not be replicated efficiently in the mutant at 34 °C (Supplementary Fig. 11). Thus, the loss of heterochromatin silencing in the *cdc20-p7* mutant seems to be coupled to a defect in DNA replication.

To gain further insight into the role of Cdc20 in the heterochromatin pathway, we analysed an amino-terminal deletion of Cdc20, denoted *cdc20<sup>AN-term</sup>*. The N terminus of Cdc20, which contains the catalytic domain, is not essential for cell survival<sup>24</sup>. To test how this mutation affects heterochromatin silencing, the *cdc20<sup>AN-term</sup>* mutant was crossed into the *otr::ade6<sup>+</sup>* background and was analysed at 23 °C on rich medium that was not supplemented with extra adenine. In this system, WT cells formed red colonies, indicating transcriptional silencing, but the *cdc20<sup>AN-term</sup>* mutant colonies were white (Fig. 2d), indicating that heterochromatin silencing is alleviated in the mutant. Consistent with this, H3K9 methylation in pericentromeric repeats was significantly reduced in the *cdc20<sup>AN-term</sup>* mutant (Fig. 2e). Thus, DNA replication and heterochromatin function were decoupled in this mutant, further showing that Cdc20 is directly involved in heterochromatin silencing.

Here we have demonstrated that the Dos2-containing complex, which includes Dos2, Mms19, Rik1 and Cdc20, is crucial for DNA replication, siRNA production and heterochromatin assembly. Our findings establish the first physical and functional link between DNA replication, small RNA generation and H3K9 methylation, and they provide a novel mechanism to explain how these processes are coordinated (Fig. 4d).

Our results provide insight into how the epigenetic states of heterochromatin are accurately duplicated in each cell cycle (Fig. 4d). In budding yeast, heterochromatin assembly requires S-phase progression but not origin firing<sup>25,26</sup>. Our findings suggest that, by contrast, DNA replication is required for heterochromatin assembly in fission yeast. In plants and mammals, DNA replication and DNA polymerase- $\epsilon$  have also been implicated in the silencing of heterochromatin<sup>27–29</sup>. This finding suggests that a molecular mechanism linking DNA replication to



**Figure 4 | Functional interactions between components of the Dos2-containing complex.** **a**, The interaction of the *cdc20-p7* mutant with Dos2 or Mms19 was abolished at 34 °C but not at 23 °C. Co-immunoprecipitations (Co-IPs) for the two sets of strains (top and bottom) incubated at the indicated temperatures were performed using an HA-tagged antibody. Precipitates were analysed by western blotting with an antibody specific for the TAP tag. **b, c**, Mms19-TAP and Dos2-TAP accumulation in heterochromatin in S phase were lost in *cdc20-p7* cells at 34 °C. ChIP assays were performed on synchronized cells carrying either Mms19-TAP or Dos2-TAP by using an antibody specific for the TAP tag, followed by PCR, as in Fig. 3c. Cell cycle progression was monitored by a septation index. The relative fold enrichment of Mms19 or Dos2 is indicated below each lane. **d**, Model of how DNA

replication, small RNA generation and H3K9 methylation are coordinated. Left, during G1/S phase, while synthesizing the leading heterochromatin strands, Cdc20 regulates heterochromatin transcription by interacting with Mms19, and it also recruits Dos2 and Rik1. Another DNA polymerase subunit may be responsible for a similar process on the lagging strand. Heterochromatin transcripts are subsequently processed into siRNAs by the RNA-induced transcriptional silencing (RITS) complex. Centre, during late S phase, the complex containing Dos1, Dos2 and Rik1, together with the siRNAs, promotes H3K9 methylation by Ctr4. Right, during G2 phase, Swi6 binds to methylated H3K9 to reassemble chromatin into a repressed state and retain cohesin.

heterochromatin formation, similar to the one elucidated in this study, is probably conserved in multicellular eukaryotes.

## METHODS SUMMARY

The *S. pombe* (fission yeast) strains used in this study are listed in Supplementary Table 1. Cell synchronization was performed by the hydroxyurea method. For mass spectrometry, TAP-tagged Dos2 was purified from a total of  $9 \times 10^{10}$  cells as described previously<sup>11</sup>. Immunofluorescence images were taken using a DeltaVision Imaging System (Applied Precision). The program SoftWoRX 2.50 (Applied Precision) was used for processing the final projections. For the ultraviolet radiation survival assay, cells were collected from the culture at logarithmic phase and were plated at appropriate dilutions onto yeast extract with supplements (YES) medium. The plates were then irradiated with various doses of ultraviolet radiation. After incubation at 23 °C for 5 days, the colonies were counted. Detailed descriptions of the co-immunoprecipitation assays, ChIP assays, RT-PCR and northern blotting are provided in the Methods.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 21 January; accepted 3 May 2011.

Published online 3 July 2011.

- Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).

- Allis, C. D., Jenuwein, T. & Reinberg, D. (eds) *Epigenetics* (Cold Spring Harbor Laboratory Press, 2006).
- Probst, A. V., Dunleavy, E. & Almouzni, G. Epigenetic inheritance during the cell cycle. *Nature Rev. Mol. Cell Biol.* **10**, 192–206 (2009).
- Wallace, J. & Orr-Weaver, T. Replication of heterochromatin: insights into mechanisms of epigenetic inheritance. *Chromosoma* **114**, 389–402 (2005).
- Chen, E. S. *et al.* Cell cycle control of centromeric repeat transcription and heterochromatin assembly. *Nature* **451**, 734–737 (2008).
- Kloc, A., Zaratiegui, M., Nora, E. & Martienssen, R. RNA interference guides histone modification during the S phase of chromosomal replication. *Curr. Biol.* **18**, 490–495 (2008).
- Motamedi, M. R. *et al.* Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**, 789–802 (2004).
- Hong, E. J., Villen, J., Gerace, E. L., Gygi, S. P. & Moazed, D. A cullin E3 ubiquitin ligase complex associates with Rik1 and the Ctr4 histone H3-K9 methyltransferase and is required for RNAi-mediated heterochromatin formation. *RNA Biol.* **2**, 106–111 (2005).
- Horn, P. J., Bastie, J. N. & Peterson, C. L. A. Rik1-associated, cullin-dependent E3 ubiquitin ligase is essential for heterochromatin formation. *Genes Dev.* **19**, 1705–1714 (2005).
- Li, F. *et al.* Two novel proteins, Dos1 and Dos2, interact with Rik1 to regulate heterochromatic RNA interference and histone modification. *Curr. Biol.* **15**, 1448–1457 (2005).
- Li, F. *et al.* Lid2 is required for coordinating H3K4 and H3K9 methylation of heterochromatin and euchromatin. *Cell* **135**, 272–283 (2008).
- Thon, G. *et al.* The Ctr7 and Ctr8 directionality factors and the Pcu4 cullin mediate heterochromatin formation in the fission yeast *Schizosaccharomyces pombe*. *Genetics* **171**, 1583–1595 (2005).
- White, S. A. & Allshire, R. C. RNAi-mediated chromatin silencing in fission yeast. *Curr. Top. Microbiol. Immunol.* **320**, 157–183 (2008).



14. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
15. D'Urso, G. & Nurse, P. *Schizosaccharomyces pombe cdc20<sup>+</sup>* encodes DNA polymerase  $\epsilon$  and is required for chromosomal replication but not for the S phase checkpoint. *Proc. Natl Acad. Sci. USA* **94**, 12491–12496 (1997).
16. Lauder, S. *et al.* Dual requirement for the yeast *MMS19* gene in DNA repair and RNA polymerase II transcription. *Mol. Cell. Biol.* **16**, 6783–6793 (1996).
17. Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
18. Hall, I. M., Noma, K. & Grewal, S. I. S. RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. *Proc. Natl Acad. Sci. USA* **100**, 193–198 (2003).
19. Jia, S. T., Noma, K. & Grewal, S. I. S. RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. *Science* **304**, 1971–1976 (2004).
20. Wu, X. Y., Li, H. & Chen, J. D. The human homologue of the yeast DNA repair and TFIIF regulator *MMS19* is an AF-1-specific coactivator of estrogen receptor. *J. Biol. Chem.* **276**, 23962–23968 (2001).
21. Ito, S. *et al.* MMD, a TFIIF-independent XPD–MMS19 protein complex involved in chromosome segregation. *Mol. Cell* **39**, 632–640 (2010).
22. Natsume, T. *et al.* A DNA polymerase  $\alpha$  accessory protein, Mcl1, is required for propagation of centromere structures in fission yeast. *PLoS ONE* **3**, e2221 (2008).
23. Nakayama, J., Allshire, R. C., Klar, A. J. S. & Grewal, S. I. S. Role for DNA polymerase  $\alpha$  in epigenetic control of transcriptional silencing in fission yeast. *EMBO J.* **20**, 2857–2866 (2001).
24. Feng, W. Y. & D'Urso, G. *Schizosaccharomyces pombe* cells lacking the amino-terminal catalytic domains of DNA polymerase  $\epsilon$  are viable but require the DNA damage checkpoint control. *Mol. Cell. Biol.* **21**, 4495–4504 (2001).
25. Kirchmaier, A. L. & Rine, J. DNA replication-independent silencing in *S. cerevisiae*. *Science* **291**, 646–650 (2001).
26. Li, Y. C., Cheng, T. H. & Gartenberg, M. R. Establishment of transcriptional silencing in the absence of DNA replication. *Science* **291**, 650–653 (2001).
27. Finnegan, E. J. & Dennis, E. S. Vernalization-induced trimethylation of histone H3 lysine 27 at FLC is not maintained in mitotically quiescent cells. *Curr. Biol.* **17**, 1978–1983 (2007).
28. Yin, H. *et al.* Epigenetic regulation, somatic homologous recombination, and abscisic acid signaling are influenced by DNA polymerase mutation in *Arabidopsis*. *Plant Cell* **21**, 386–402 (2009).
29. Fuss, J. & Linn, S. Human DNA polymerase  $\epsilon$  colocalizes with proliferating cell nuclear antigen and DNA replication late, but not early, in S phase. *J. Biol. Chem.* **277**, 8658–8666 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank R. Allshire, P. Nurse, G. D'Urso and the Japan Yeast Genetic Resource Center for strains, E. Osborne and C. Hale for comments on the manuscript, Cold Spring Harbor Laboratory for mass spectrometry analysis, and members of the Cande and Li laboratories for their support and discussions. This work was supported by a grant from the National Institutes of Health (R01GM076396) to W.Z.C. and R.M.

**Author Contributions** F.L. and W.Z.C. designed the experiments and wrote the manuscript. F.L. performed the experiments. R.M. provided mass spectrometry expertise, equipment and conceptual support.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to F.L. (fl43@nyu.edu).

## METHODS

**Yeast strains, media and genetic procedures.** The *S. pombe* (fission yeast) strains used in this study are listed in Supplementary Table 1. Yeast extract with supplements (YES) was used as a complete culture medium, Edinburgh minimal medium (EMM) as a minimal medium, and SPAS medium for conjugation and sporulation. Cell synchronization was performed by the hydroxyurea method. Briefly, cells were treated with 12 mM hydroxyurea for 4 h and then released into 100  $\mu\text{g ml}^{-1}$  thiabendazole for 1 h to block entry to mitosis. Standard genetic protocols for fission yeast were used<sup>30</sup>.

**Mass spectrometry.** TAP-tagged Dos2 was purified from a total of  $9 \times 10^{10}$  cells as described previously<sup>11</sup>. Briefly, cell lysates in 1 $\times$  lysis buffer (50 mM bis-Tris propane, pH 7.0, 0.1 M KCl, 5 mM EDTA, 5 mM EGTA and 10% glycerol) were incubated with IgG sepharose (Amersham Pharmacia Biotech) for 2 h. After washing with lysis buffer, the IgG sepharose was incubated overnight with TEV protease (Invitrogen). The supernatant was removed from the IgG sepharose and added to S-protein agarose slurry (Novagen) for 3 h. The S-protein agarose was then washed with lysis buffer. The eluate from this agarose was analysed by silver staining and subjected to mass spectrometry (at Cold Spring Harbor Laboratory).

**Immunoprecipitation assays.** Cells were lysed in HB buffer by using the glass bead method<sup>30</sup>. Lysates were pre-cleared with protein A agarose beads, which was followed by a 2-h incubation with anti-HA or anti-GFP antibody (Sigma) or IgG Sepharose 6 Fast Flow beads (Amersham Biosciences) at 4 °C. After washing, the eluted proteins and input extracts were analysed by western blotting using anti-S-tag (MA1-981, ABR Affinity BioReagents) or anti-RNA Pol II (ab5408, Abcam) antibody.

**ChIP analysis.** CHIP assays were carried out as described previously<sup>31</sup>. Cells taken from culture at logarithmic phase were crosslinked with 1% formaldehyde. Immunoprecipitation was performed with S-protein agarose (Novagen) or antibodies specific for the following: dimethylated H3K9 (07-441, Upstate) or Swi6 (ab14898, Abcam). The precipitated DNA was analysed by using competitive PCR with oligonucleotides specific for the centromeric *dh* region or a control gene, *act1*<sup>+</sup>. The PCR products were separated on a 1.7% agarose gel and post-stained with ethidium bromide. The primers used are listed in Supplementary Table 2.

**RT-PCR.** Total RNA was isolated from cells taken from culture growing at logarithmic phase by using an RNeasy Mini Kit (Qiagen). After treatment with DNase I (Promega), 50 ng purified RNA was analysed by RT-PCR in a 25  $\mu\text{l}$

reaction volume using a one-step RT-PCR kit (Qiagen). Equal loading of RNA samples was assessed by amplification of a control gene, *act1*<sup>+</sup>. For strand-specific RT-PCR, RNA samples were incubated with primers that were complementary to the forward or the reverse centromeric transcripts for synthesis of the first cDNA strand. After heat inactivation of the reverse transcriptase at 95 °C for 15 min, a second primer was added for the subsequent cycles of PCR amplification. The primers used for RT-PCR are listed in Supplementary Table 2.

**Small RNA northern blotting analyses.** Small RNA northern blotting analysis was performed as described previously<sup>10</sup>. Briefly, siRNAs were extracted from cells taken when the culture was in logarithmic phase, in YES medium, using a mirVana miRNA Isolation Kit (Ambion). Total small RNA (25  $\mu\text{g}$ ) was resolved on a 15% denaturing acrylamide gel and was blotted to a charged nylon membrane (Hybond-N+, Amersham). RNA blots were crosslinked and hybridized with DNA probes specific for the centromeric *otr* region or snoR69 as a loading control. The DNA oligonucleotides used as probes are listed in Supplementary Table 2.

**Microscopy.** Fluorescent images were taken using a DeltaVision Imaging System (Applied Precision). The program SoftWoRX 2.50 (Applied Precision) was used for processing the final projections. The standard procedure was followed by staining with 4',6-diamidino-2-phenylindole (DAPI).

**Ultraviolet radiation survival assay.** Cells were collected when the culture was in logarithmic phase and were plated at appropriate dilutions onto YES medium. The plates were irradiated with various doses of ultraviolet radiation. After incubation at 23 °C for 5 days, the colonies were counted.

**DNA replication assay.** A pBluescript plasmid carrying a 3.0-kb fragment from heterochromatic *cen3* repeats (3.0-K)<sup>32</sup>, together with a *ura4*<sup>+</sup> selective marker, was transformed into the *cdc20-p7* mutant. This centromeric fragment contains efficient ARS elements<sup>32</sup>. After incubation at 23 °C or 34 °C for 4 days, total DNA was isolated following standard procedures<sup>32</sup>. Purified DNA (50 ng) was analysed by PCR (Bio-Rad) with primers specific for *ura4*<sup>+</sup>. Equal loading was assessed by amplification of a control gene, *act1*<sup>+</sup>.

30. Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823 (1991).

31. Pidoux, A., Mellone, B. & Allshire, R. Analysis of chromatin in fission yeast. *Methods* **33**, 252–259 (2004).

32. Smith, J. G. *et al.* Replication of centromere II of *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* **15**, 5165–5172 (1995).

# Subunit arrangement and phenylethanolamine binding in GluN1/GluN2B NMDA receptors

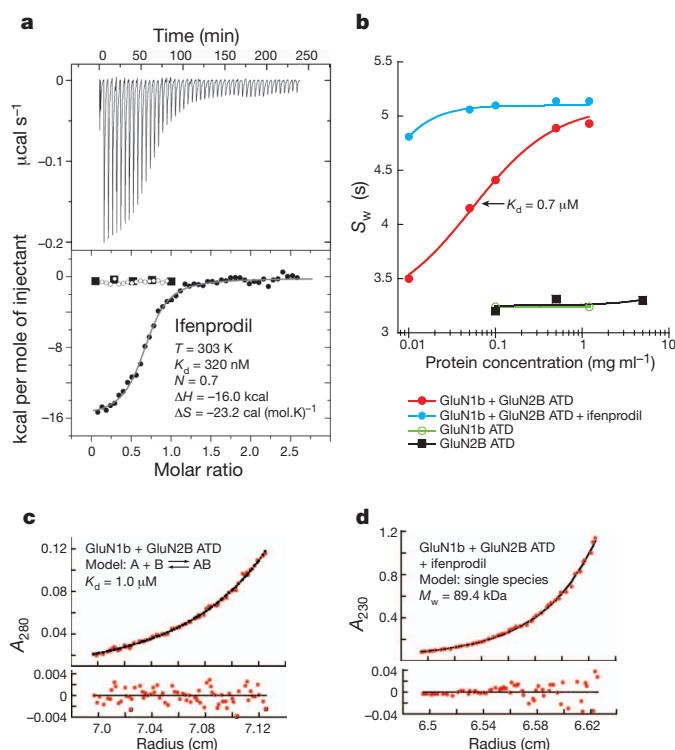
Erkan Karakas<sup>1</sup>, Noriko Simorowski<sup>1</sup> & Hiro Furukawa<sup>1</sup>

Since it was discovered that the anti-hypertensive agent ifenprodil has neuroprotective activity through its effects on NMDA (N-methyl-D-aspartate) receptors<sup>1</sup>, a determined effort has been made to understand the mechanism of action and to develop improved therapeutic compounds on the basis of this knowledge<sup>2–4</sup>. Neurotransmission mediated by NMDA receptors is essential for basic brain development and function<sup>5</sup>. These receptors form heteromeric ion channels and become activated after concurrent binding of glycine and glutamate to the GluN1 and GluN2 subunits, respectively. A functional hallmark of NMDA receptors is that their ion-channel activity is allosterically regulated by binding of small compounds to the amino-terminal domain (ATD) in a subtype-specific manner. Ifenprodil and related phenylethanolamine compounds, which specifically inhibit GluN1 and GluN2B NMDA receptors<sup>6,7</sup>, have been intensely studied for their potential use in the treatment of various neurological disorders and diseases, including depression, Alzheimer's disease and Parkinson's disease<sup>2,4</sup>. Despite considerable enthusiasm, mechanisms underlying the recognition of phenylethanolamines and ATD-mediated allosteric inhibition remain limited owing to a lack of structural information. Here we report that the GluN1 and GluN2B ATDs form a heterodimer and that phenylethanolamine binds at the interface between GluN1 and GluN2B, rather than within the GluN2B cleft. The crystal structure of the heterodimer formed between the GluN1b ATD from *Xenopus laevis* and the GluN2B ATD from *Rattus norvegicus* shows a highly distinct pattern of subunit arrangement that is different from the arrangements observed in homodimeric non-NMDA receptors and reveals the molecular determinants for phenylethanolamine binding. Restriction of domain movement in the bi-lobed structure of the GluN2B ATD, by engineering of an inter-subunit disulphide bond, markedly decreases sensitivity to ifenprodil, indicating that conformational freedom in the GluN2B ATD is essential for ifenprodil-mediated allosteric inhibition of NMDA receptors. These findings pave the way for improving the design of subtype-specific compounds with therapeutic value for neurological disorders and diseases.

The consensus view that has emerged from functional studies of NMDA receptors using site-directed mutagenesis and molecular modelling is that phenylethanolamine compounds such as ifenprodil and Ro 25-6981 bind to the ATD of the GluN2B subunit. However, this has not been established directly and the mechanism of action is complicated by the obligate heteromeric assembly of NMDA receptors. To establish directly that phenylethanolamines bind to the ATDs of these receptors, we used isothermal titration calorimetry to measure the binding of ifenprodil and Ro 25-6981 to purified recombinant *Rattus norvegicus* GluN2B (residues 31–394) and *Xenopus laevis* GluN1b (residues 23–408) ATDs (Supplementary Fig. 1). GluN1b from *Xenopus laevis*<sup>8,9</sup> was used in this study because of its superior biochemical stability compared to other orthologues. It is 93% identical in primary sequence to the *Rattus norvegicus* GluN1 ATD and is capable of forming functional NMDA-receptor ion channels that undergo ifenprodil inhibition when combined with *Rattus norvegicus* GluN2B<sup>9</sup> (Supplementary Fig. 2).

When the GluN1b ATD or GluN2B ATD proteins were individually injected with ifenprodil, there was no evidence of binding (Fig. 1a). However, when a mixture of the GluN1b and GluN2B ATD proteins was injected with ifenprodil or Ro 25-6981, a dose-dependent heat exchange was observed, with dissociation constant ( $K_d$ ) values of 320 nM and 60 nM, respectively (Fig. 1a and Supplementary Fig. 3). Thus, both the GluN1b and GluN2B ATDs are required for binding of phenylethanolamines.

The necessity of both ATDs for recognition of phenylethanolamine indicates that binding takes place in the GluN1–GluN2B heteromer. To probe the association pattern of GluN1b and GluN2B ATD proteins,



**Figure 1 | Binding of phenylethanolamine requires both GluN1b and GluN2B ATDs, and stabilizes heterodimers.** **a**, Calorimetric titration of ifenprodil into a GluN1b and GluN2B ATD mixture (upper panel) and integrated heat as a function of the ifenprodil/protein molar ratio (lower panel) for GluN1b ATD (open circles), GluN2B ATD (filled squares) and the GluN1b/GluN2B ATD mixture (filled circles). **b**, Weighted-average sedimentation coefficient ( $S_w$ ) for GluN1b ATD alone (green), GluN2B ATD alone (black) and the GluN1b–GluN2B ATD mixture in the presence (cyan) and absence (red) of 10  $\mu$ M ifenprodil, fitted with a monomer-dimer model (lines). **c**, **d**, Sedimentation equilibrium analysis of GluN1b and GluN2B ATDs in the absence (**c**) and presence (**d**) of 10  $\mu$ M ifenprodil. Data points at a rotor speed of 18,000 r.p.m. (red dots) are shown with a global fit (black line) of the data. Residuals from the fit are shown in the lower panel.

<sup>1</sup>Cold Spring Harbor Laboratory, WM Keck Structural Biology Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.

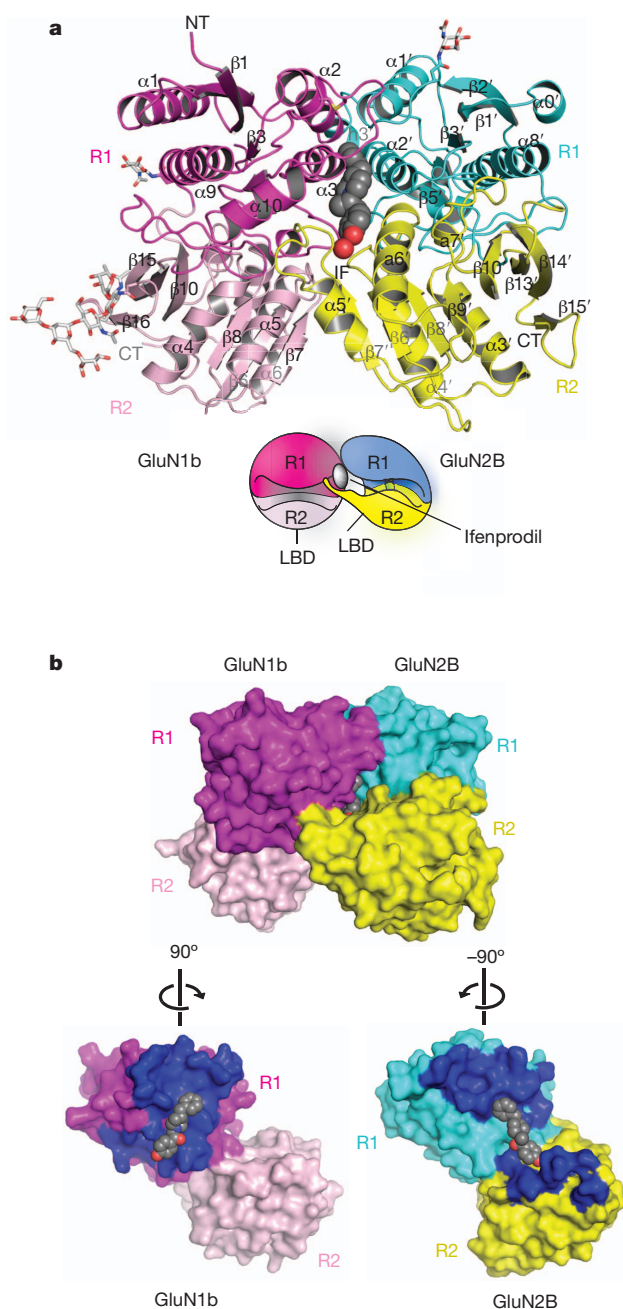


we determined the mass of the ATD proteins in solution by sedimentation experiments (Fig. 1b–d). Although the individual GluN1b ATD and GluN2B ATD were exclusively monomeric at  $1.2 \text{ mg ml}^{-1}$  (Fig. 1b), they formed a heterodimer with a  $K_d$  of  $0.7\text{--}1 \mu\text{M}$  when mixed together (Fig. 1b, c). Notably, when ifenprodil was included in the GluN1b/GluN2B ATD protein mixture, the heterodimerization was strengthened by at least 20-fold (Fig. 1b, d). These results establish that the GluN1b and GluN2B ATDs form heterodimers and that phenylethanolamines probably bind at the GluN1b–GluN2B subunit interface.

To understand the nature of the subunit interaction between GluN1b and GluN2B at their ATDs, and to pinpoint the location of the phenylethanolamine binding site, we conducted crystallographic studies on the GluN1b and GluN2B ATD proteins (Supplementary Table 1). The crystallographic analysis showed that the GluN1b and GluN2B ATDs exist as heterodimers in both ifenprodil-bound and Ro 25-6981-bound forms (Fig. 2). No notable structural difference was observed between the monomers of GluN1b ATD (Supplementary Fig. 4) or GluN2B ATD<sup>10</sup> and the respective subunits in the GluN1b–GluN2B ATD complex, indicating that dimerization did not cause changes in the overall conformation. Most notably, the crystal structures clearly identified the phenylethanolamine binding site at the heterodimer interface (Fig. 2).

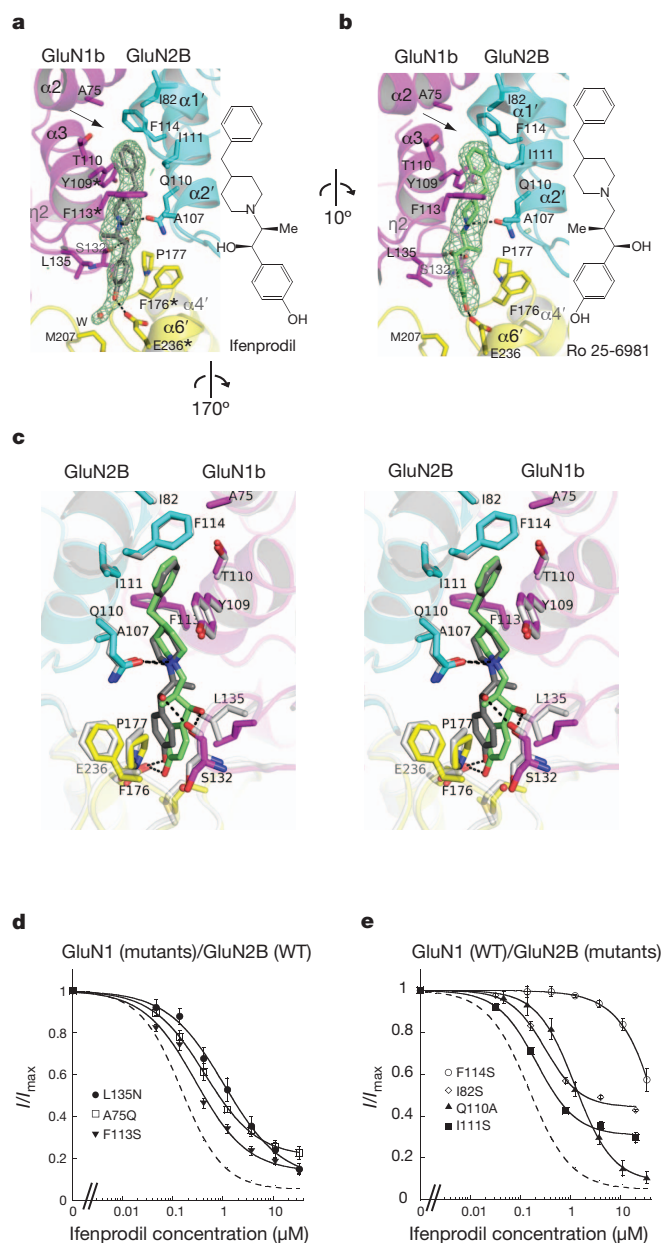
Both the GluN1b and GluN2B ATDs have bi-lobed clamshell-like architectures composed of R1 and R2 domains that are roughly similar in secondary-structure distribution to non-NMDA-receptor ATDs<sup>11–14</sup>. However, the structures of the GluN1b and GluN2B ATD monomers cannot be superimposed onto non-NMDA-receptor ATD monomers, owing to a major difference in the R1–R2 orientations, as was also observed previously in a study of the GluN2B ATD monomer<sup>10</sup> (Supplementary Fig. 5). The unique R1–R2 orientations of the GluN1b and GluN2B ATDs result in a heterodimer assembly that is distinct from that observed in non-NMDA-receptor ATD homodimers<sup>11–14</sup> (Supplementary Fig. 6). Whereas non-NMDA-receptor ATD subunits form symmetrical homodimers through strong R1–R1 and R2–R2 interactions, the GluN1b and GluN2B ATDs associate with each other asymmetrically through R1–R1 and R1(GluN1b)–R2(GluN2B) interactions<sup>11,12</sup> (Fig. 2). No residue from GluN1b R2 is involved in the GluN1b–GluN2B interaction (Fig. 2b). The R1–R1 interface contains hydrophobic interactions mediated by residues from the cores of the  $\alpha 2$  helix and  $\alpha 3$  helix in GluN1b, and from the  $\alpha 1'$  helix and  $\alpha 2'$  helix in GluN2B, surrounded by polar interactions involving the GluN1b  $\alpha 2$  helix, the GluN2B  $\alpha 1'$  helix and the hypervariable loops<sup>10</sup> (Supplementary Fig. 7). The R1–R2 interface involves mainly polar interactions, involving residues on the  $\alpha 10$  helix, a loop extending from  $\eta 2$  in GluN1b and loops extending from the  $\beta 6'$  sheet and  $\beta 7'$  sheet in GluN2B (Fig. 2d). The lack of R2–R2 interaction in the GluN1b and GluN2B ATDs leaves sufficient room for the previously suggested conformational movement of the bi-lobed structure in GluN2B<sup>10,15</sup>, which is important in mediating the allosteric regulation that is unique to NMDA receptors. In non-NMDA receptors, such movement is prohibited, owing to strong R2–R2 interactions that lock the movement of R2 (refs 3, 11–13).

The heterodimeric arrangement of GluN1b and GluN2B creates a phenylethanolamine binding pocket composed of residues from GluN1b R1, GluN2B R1 and GluN2B R2 (Fig. 2). The phenylethanolamine binding site has no overlap with the zinc binding site that is located in the GluN2B ATD cleft<sup>10,16</sup> (Supplementary Fig. 8). In the crystal structure, ifenprodil is buried in the dimer interface with insufficient space for entering or exiting (Fig. 2b), which indicates that binding occurs through an induced-fit mechanism and that unbinding may involve opening of the GluN2B ATD bi-lobed structure. All of the residues at the binding sites are identical among *Xenopus laevis*, rat and human orthologues, indicating that inhibition of NMDA receptors by phenylethanolamine is a conserved feature among those species (Supplementary Fig. 9). Binding of both ifenprodil and Ro 25-6981 is



**Figure 2 | Structure of the GluN1b–GluN2B ATD heterodimer in complex with ifenprodil at 2.6 Å resolution.** **a**, View of the ATD heterodimer from the side. The GluN1b and GluN2B ATDs have bi-lobed architecture composed of R1 (magenta and cyan) and R2 (light pink and yellow) domains. Ifenprodil (grey spheres) sits at the heterodimer interface. N-glycosylation chains are shown in white. NT, N terminus; CT, C terminus. The cartoon shows an approximate orientation of the GluN1b and GluN2B ATDs with black sticks below R2 indicating the C-terminal ends where ligand-binding domains (LBDs) begin. **b**, Surface presentation of the GluN1b–GluN2B ATD heterodimer (upper panel) and of each subunit (lower panel), showing residues at the subunit interface in dark blue. Note that ifenprodil (grey spheres) is occluded in the subunit interface. The heterodimer buries  $1,191 \text{ Å}^2$  of solvent-accessible surface area per subunit, with the GluN1b R1–GluN2B R1 and GluN1b R1–GluN2B R2 interfaces contributing 62% and 38%, respectively.

mediated primarily through hydrophobic interactions between the benzylpiperidine group and a cluster of hydrophobic residues from the GluN1b  $\alpha 2$  helix and  $\alpha 3$  helix and the GluN2B  $\alpha 1'$  helix and  $\alpha 2'$  helix, and between the hydroxylphenyl groups and GluN1b Leu 135, GluN2B Phe 176 and GluN2B Pro 177 (Fig. 3a, b). Furthermore, the



**Figure 3 | Phenylethanolamine binding site.** **a, b**, Binding of ifenprodil (**a**) and Ro 25-6981 (**b**) takes place at the GluN1b–GluN2B subunit interface. Mesh represents the  $F_o - F_c$  omit electron density map contoured at  $3\sigma$ . Residues marked with asterisks in **a** have been previously shown to affect ifenprodil sensitivity. Adjacent to the binding pocket is an empty space surrounded by hydrophobic residues, including GluN1b Ala 75, GluN2B Ile 82 and GluN2B Phe 114 (arrows). **c**, Comparison of binding patterns of ifenprodil (grey) and Ro 25-6981 (lime) in stereoview. The structure bound to Ro 25-6981 is coloured as in **b**, whereas the ifenprodil-bound structure is coloured white. **d, e**, New residues found to interact with phenylethanolamines in this study were mutated and analysed for their effect on sensitivity to ifenprodil. Mutation of the residues surrounding the binding site caused changes in  $IC_{50}$  as well as changes in the extent of inhibition by ifenprodil. WT, wild-type;  $I/I_{max}$  relative current with ( $I$ ) and without ( $I_{max}$ ) ifenprodil. Error bars represent s.d.

drugs make three direct polar interactions with Ser 132 of GluN1b, Glu 110 of GluN2B and Asp 236 of GluN2B. Superposition of the binding sites of ifenprodil and Ro 25-6981 shows that the methyl and hydroxyl groups in the propanol moiety of both ligands face in opposite directions and that the benzylpiperidine groups sit in the binding pocket in similar ways (Fig. 3c). Consequently, Ro 25-6981 has a higher affinity for GluN1/GluN2B NMDA receptors than ifenprodil<sup>17</sup> because the methyl group in Ro 25-6981 is in a favourable

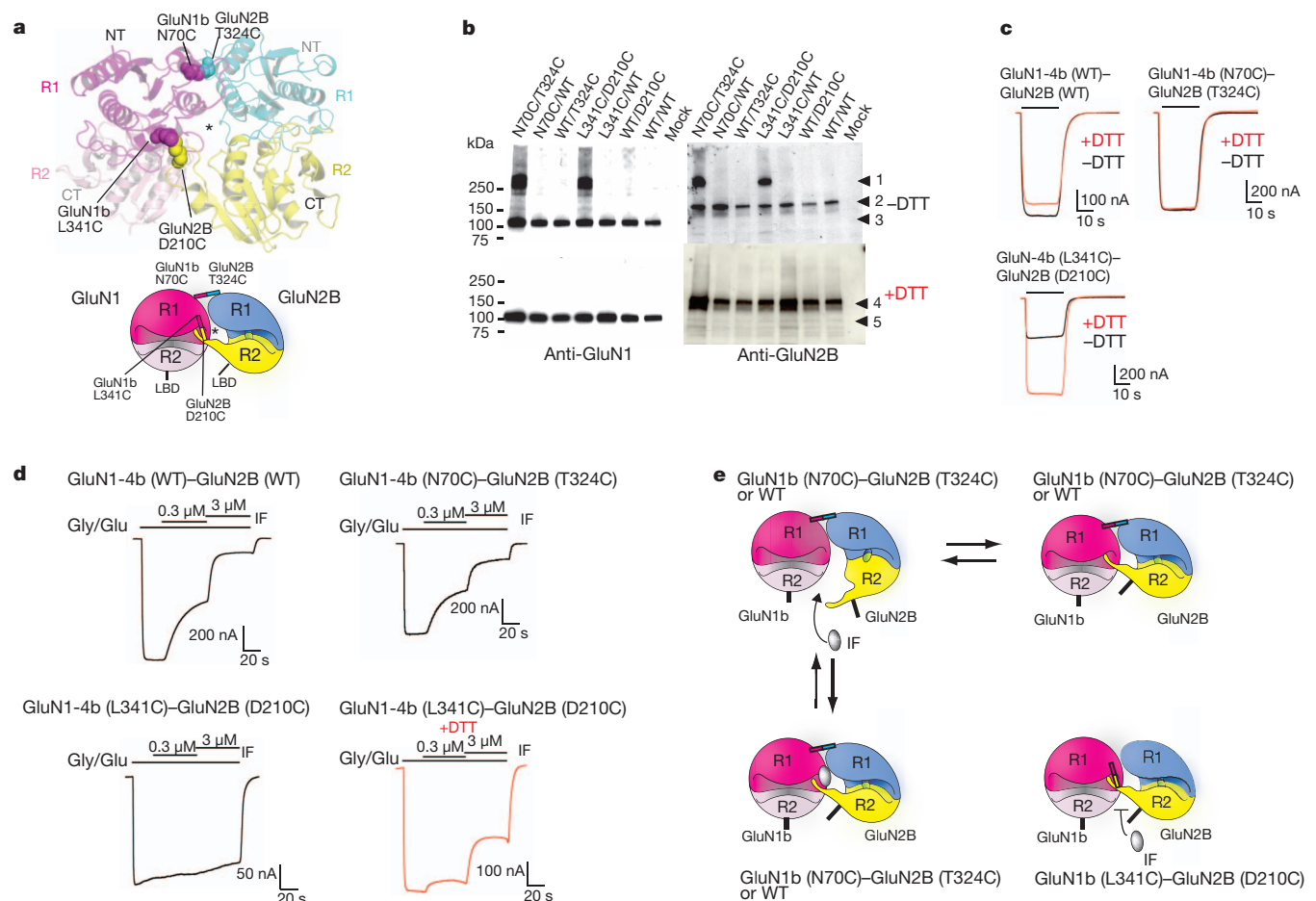
position to form a hydrophobic interaction involving Phe 176 and Pro 177 in the GluN2B subunit, whereas ifenprodil makes a weaker hydrophobic interaction with GluN1b, involving Leu 135. Extensive mutagenesis studies have previously indicated that GluN1b Tyr 109 (ref. 18) and GluN2B Phe 176 and Asp 236 (ref. 19) are critical in mediating inhibition by ifenprodil, but whether these are involved in binding or transducing the inhibitory effect was unknown. We performed additional mutagenesis studies on newly identified residues in both GluN1b and GluN2B at the ifenprodil binding site, measured macroscopic currents by two-electrode voltage clamp, and revealed significant alterations in  $IC_{50}$  and in the extent of inhibition (Fig. 3d, e and Supplementary Table 2), thereby confirming the physiological relevance of the binding site. Notably, disruption of the 'empty' hydrophobic space formed by GluN1b Ala 75, GluN2B Ile 82 and GluN2B Phe 114 (arrows in Fig. 3a and b) by site-directed mutations to hydrophilic residues had marked effects on sensitivity to ifenprodil (Fig. 3d, e). Thus, stabilization of this hydrophobic space by filling it with a hydrophobic moiety may be a valid strategy to improve the design of phenylethanolamine-based drugs.

It is not known why phenylethanolamine binds specifically to the GluN1–GluN2B subunit combination. Although inspection of the primary sequences shows non-conservation of the critical binding-site residues between GluN2B and GluN2C or GluN2D (for example, the equivalent residue to GluN2B Phe 176 is not conserved in GluN2C or GluN2D), all of the residues in GluN2A are conserved except for GluN2B Ile 111 (Met 112 in GluN2A) (Supplementary Fig. 10). Indeed, the mutations GluN2A Met112Ile or GluN2B Ile111Met do not confer or abolish ifenprodil sensitivity in GluN1/GluN2A or GluN1/GluN2B receptors, respectively (Supplementary Table 2). Thus, the insensitivity of the GluN1/GluN2A receptors to phenylethanolamine may stem from a fundamental difference in the mode of subunit association between GluN1/GluN2A and GluN1/GluN2B at their ATDs.

To validate further the physiological relevance of the heterodimeric assembly, we engineered cysteine mutants at the subunit interface, using the ifenprodil-bound GluN1b/GluN2B ATD structure as a guide, in the context of the intact rat GluN1-4b/GluN2B receptor. These cysteines were designed to form spontaneous disulphide bonds if the mutated residues were proximal to each other. We designed two pairs of cysteine mutants, GluN1-4b (Asn70Cys) with GluN2B (Thr324Cys), and GluN1-4b (Leu341Cys) with GluN2B (Asp210Cys). These mutations 'lock' the R1–R1 and R1–R2 interfaces, respectively (Fig. 4a). We then expressed the mutant receptors in mammalian cell cultures and analysed them for formation of disulphide-linked oligomers in western blots. When mutant receptors of one subunit were co-expressed with wild-type receptors of the other, they gave rise to monomeric bands that were identical to wild-type GluN1-4b–GluN2B receptors in both reducing and non-reducing conditions (110 kDa and 170 kDa for GluN1-4b and GluN2B, respectively; Fig. 4b, arrows 2 and 3). In contrast, co-expressing pairs of the GluN1-4b–GluN2B cysteine mutants gave rise to a heterodimeric ~280 kDa band that was recognized by both anti-GluN1 and anti-GluN2B antibodies in non-reducing conditions (Fig. 4b, arrow 1). This confirms that the R1–R1 and R1–R2 subunit interfaces observed in the GluN1b–GluN2B ATD crystal structures are physiological and that the heterodimer, not the homodimer, is the basic functional unit in the ATD of the NMDA receptor<sup>20</sup>. Furthermore, disulphide crosslinking was observed in the presence and absence of ifenprodil, indicating that the ligand-free GluN1b–GluN2B ATDs may oscillate between the previously suggested open conformation<sup>15</sup> and the closed conformation represented by the crystal structure described here.

To understand the functional effects of locking the R1–R1 and R1–R2 interactions in the GluN1b and GluN2B ATDs, we measured macroscopic current responses from the ion channels of the cysteine-mutant receptors by two-electrode voltage clamp. First, we explored the effect on ion-channel activity of breaking the disulphide bonds.





**Figure 4 | Engineering of disulphide bonds at the subunit interface alters sensitivity to ifenprodil.** **a**, Location of mutated residues at the R1–R1 and R1–R2 interfaces in the GluN1b and GluN2B ATDs (spheres), and location of the ifenprodil binding pocket (asterisk). **b**, Detection of disulphide bonds by anti-GluN1 and anti-GluN2B western blots in reducing (+DTT) and non-reducing (–DTT) conditions. Arrow 1, GluN1-4b–GluN2B heterodimer; arrows 2 and 4, GluN2B monomers; arrows 3 and 5, GluN1-4b monomers. **c**, Macroscopic current recording of the wild-type and mutant receptors in the presence (red)

and absence (black) of DTT (2 mM). **d**, Effect of disulphide bonds on the sensitivity to ifenprodil (IF) of wild-type and mutant receptors in the presence (red) and absence (black) of DTT. **e**, Possible model of ifenprodil binding and the movement of ATDs for allosteric inhibition. Ifenprodil binds to the open GluN2B clamshell and induces domain closure, resulting in allosteric inhibition. In the GluN1-4b (Asn70Cys)–GluN2B (Thr324Cys) receptor, the GluN2B ATD is locked in the closed conformation so ifenprodil cannot access the binding site.

Application of dithiothreitol (DTT) had a minor inhibitory effect on wild-type GluN1-4b–GluN2B receptors and on receptors containing GluN1-4b (Asn70Cys) and GluN2B (Thr324Cys). In contrast, a 2.5-fold potentiation was observed on breakage of the disulphide bond at the R1–R2 interface between GluN1-4b (Leu341Cys) and GluN2B (Asp210Cys) (Fig. 4c and Supplementary Fig. 11). This implies that locking the closed conformation in the GluN2B ATD bi-lobed structure by the R1–R2 crosslink results in downregulation of ion-channel activity. We next tested the effects of the disulphide bonds on sensitivity to ifenprodil. Although the R1–R1 crosslink had only a minor effect, the R1–R2 crosslink almost completely abolished inhibition by ifenprodil, even at 3  $\mu$ M (Fig. 4d). When this R1–R2 disulphide crosslink was broken by the application of DTT, the mutant receptors regained sensitivity to ifenprodil, to a similar extent to that of receptors composed of wild-type GluN1-4b and GluN2B (Asp210Cys) in non-reducing conditions (Fig. 4d and Supplementary Fig. 12). This indicates that ifenprodil cannot bind to the GluN1b–GluN2B ATD when the R1–R2 interaction is locked and thus, when the GluN2B ATD clamshell is closed. Taken together, the experiments described above indicate that the binding of ifenprodil requires an opening of the GluN2B bi-lobed structure and that inhibition by ifenprodil involves closure of the clamshell through the GluN1b R1–GluN2B R2 interaction (Fig. 4e).

This study shows that phenylethanolamine binds at the GluN1–GluN2B subunit interface through an induced-fit mechanism and that allosteric inhibition involves stabilization of the GluN2B ATD clamshell structure in a closed conformation. The binding mechanism presented here provides a molecular blueprint for improving the design of therapeutic compounds targeting the ATD of the NMDA receptor.

## METHODS SUMMARY

GluN1b and GluN2B ATDs were expressed as secreted proteins using the insect-cell/baculovirus system and purified using metal-chelate chromatography and size-exclusion chromatography. Crystallization was performed in hanging-drop vapour diffusion configuration in a buffer containing 20% PEG3350, 150 mM KNO<sub>3</sub> and 50 mM HEPES–NaOH (pH 7.0) for the GluN1b ATD, or 3.0–3.5 M sodium formate and 0.1 M HEPES–NaOH (pH 7.5) for the GluN1b–GluN2B ATD heterodimer. Diffraction data sets obtained at 100 K were indexed, integrated and scaled using HKL2000. The GluN1b ATD structure was solved by the single anomalous diffraction phasing method using Se–Met-incorporated crystals, and the GluN1b–GluN2B ATD structures were solved by molecular replacement using coordinates of GluN1b ATD and GluN2B ATD (Protein Data Bank code 3JPW<sup>10</sup>). Model refinement was conducted using the program Phenix<sup>21</sup>. Experiments involving analytical ultracentrifugation and isothermal titration calorimetry were conducted using the purified protein samples in their glycosylated form. Ion-channel activities of full-length NMDA receptors were measured by whole-cell recording



from cRNA-injected *Xenopus laevis* oocytes, using a two-electrode voltage-clamp configuration.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 20 January; accepted 6 May 2011.**

**Published online 15 June 2011.**

- Gotti, B. *et al.* Ifenprodil and SL 82.0715 as cerebral anti-ischemic agents. I. Evidence for efficacy in models of focal cerebral ischemia. *J. Pharmacol. Exp. Ther.* **247**, 1211–1221 (1988).
- Koller, M. & Urwyler, S. Novel *N*-methyl-D-aspartate receptor antagonists: a review of compounds patented since 2006. *Expert Opin. Ther. Pat.* **20**, 1683–1702 (2010).
- Hansen, K. B., Furukawa, H. & Traynelis, S. F. Control of assembly and function of glutamate receptors by the amino-terminal domain. *Mol. Pharmacol.* **78**, 535–549 (2010).
- Mony, L., Kew, J. N., Gunthorpe, M. J. & Paoletti, P. Allosteric modulators of NR2B-containing NMDA receptors: molecular mechanisms and therapeutic potential. *Br. J. Pharmacol.* **157**, 1301–1317 (2009).
- Traynelis, S. F. *et al.* Glutamate receptor ion channels: structure, regulation, and function. *Pharmacol. Rev.* **62**, 405–496 (2010).
- Gallagher, M. J., Huang, H., Pritchett, D. B. & Lynch, D. R. Interactions between ifenprodil and the NR2B subunit of the *N*-methyl-D-aspartate receptor. *J. Biol. Chem.* **271**, 9603–9611 (1996).
- Williams, K. Ifenprodil discriminates subtypes of the *N*-methyl-D-aspartate receptor: selectivity and mechanisms at recombinant heteromeric receptors. *Mol. Pharmacol.* **44**, 851–859 (1993).
- Ewald, R. C. & Cline, H. T. Cloning and phylogenetic analysis of NMDA receptor subunits NR1, NR2A and NR2B in *Xenopus laevis* tadpoles. *Front. Mol. Neurosci.* **2**, 4 (2009).
- Schmidt, C. & Hollmann, M. Molecular and functional characterization of *Xenopus laevis* *N*-methyl-D-aspartate receptors. *Mol. Cell. Neurosci.* **42**, 116–127 (2009).
- Karakas, E., Simorowski, N. & Furukawa, H. Structure of the zinc-bound amino-terminal domain of the NMDA receptor NR2B subunit. *EMBO J.* **28**, 3910–3920 (2009).
- Kumar, J., Schuck, P., Jin, R. & Mayer, M. L. The N-terminal domain of GluR6-subtype glutamate receptor ion channels. *Nature Struct. Mol. Biol.* **16**, 631–638 (2009).
- Jin, R. *et al.* Crystal structure and association behaviour of the GluR2 amino-terminal domain. *EMBO J.* **28**, 1812–1823 (2009).
- Clayton, A. *et al.* Crystal structure of the GluR2 amino-terminal domain provides insights into the architecture and assembly of ionotropic glutamate receptors. *J. Mol. Biol.* **392**, 1125–1132 (2009).
- Sobolevsky, A. I., Rosconi, M. P. & Gouaux, E. X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature* **462**, 745–756 (2009).
- Gielen, M., Siegler Retchless, B., Mony, L., Johnson, J. W. & Paoletti, P. Mechanism of differential control of NMDA receptor activity by NR2 subunits. *Nature* **459**, 703–707 (2009).
- Rachline, J., Perin-Dureau, F., Le Goff, A., Neyton, J. & Paoletti, P. The micromolar zinc-binding domain on the NMDA receptor subunit NR2B. *J. Neurosci.* **25**, 308–317 (2005).
- Malherbe, P. *et al.* Identification of critical residues in the amino terminal domain of the human NR2B subunit involved in the RO 25-6981 binding pocket. *J. Pharmacol. Exp. Ther.* **307**, 897–905 (2003).
- Masuko, T. *et al.* A regulatory domain (R1–R2) in the amino terminus of the *N*-methyl-D-aspartate receptor: effects of spermine, protons, and ifenprodil, and structural similarity to bacterial leucine/isoleucine/valine binding protein. *Mol. Pharmacol.* **55**, 957–969 (1999).
- Perin-Dureau, F., Rachline, J., Neyton, J. & Paoletti, P. Mapping the binding site of the neuroprotectant ifenprodil on NMDA receptors. *J. Neurosci.* **22**, 5955–5965 (2002).
- Lee, C. H. & Gouaux, E. Amino terminal domains of the NMDA receptor are organized as local heterodimers. *PLoS ONE* **6**, e19181 (2011).
- Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the staff at X25 and X29 at the National Synchrotron Light Source for beamline support. M. Mayer is thanked for comments on this work. We also thank D. Raleigh for the use of analytical ultracentrifugation. GluN1 clones from *Xenopus laevis* were gifts from M. Hollmann and H. Cline. This work was supported by NIH MH085926, the Alzheimer's Association and a donation from the Fox family (to H.F.). H.F. was also funded by a scientist development grant from the American Heart Association. E.K. is supported by a NARSAD Lieber Young Investigator Award.

**Author Contributions** The project was initiated by E.K. and H.F. All of the experiments were designed by E.K. and H.F. Crystallographic studies, isothermal calorimetry and analytical ultracentrifugation were carried out by E.K. Electrophysiology and crosslinking experiments were conducted by H.F. Technical support was given by N.S. The manuscript was written by H.F. and E.K.

**Author Information** Structural coordinates are deposited in the Protein Data Bank with accession codes 3QEK for GluN1b-ATD, 3QEL for GluN1b–GluN2B-ATD in complex with ifenprodil and 3QEM for GluN1b–GluN2B-ATD in complex with Ro 25-6981. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to H.F. ([furukawa@cshl.edu](mailto:furukawa@cshl.edu)).

## METHODS

**Expression, purification and crystallization of GluN1b and GluN2B ATDs.**

The *Xenopus laevis* GluN1b ATD (Met1 to Glu408), containing Cys22Ser, Asn61Gln and Asn371Gln mutations, was C-terminally fused to a thrombin cleavage site followed by an octa-histidine tag. The *Xenopus laevis* GluN1b ATD and rat GluN2B ATD<sup>10</sup> constructs were individually expressed or co-expressed using the High Five (*Trichoplusia ni*) baculovirus system (DH10multibac)<sup>22</sup>. Purification was performed using a similar method to that described previously<sup>10</sup> except that the proteins were de-glycosylated by endoglycosidase F1 after purification by metal-chelate chromatography, and 1  $\mu$ M ifenprodil or 1  $\mu$ M Ro 25-6981 was included in the running buffer of size-exclusion chromatography (Superdex200) for isolation of the GluN1b–GluN2B ATD complex. The proteins used for isothermal titration calorimetry and sedimentation experiments were purified without the endoF1 de-glycosylation step and in the absence of ifenprodil or Ro 25-6981. Se-Met-incorporated GluN1b ATD proteins were expressed using methionine-free media (ESF921) supplemented with DL-Se-Met (Sigma) at 100 mg l<sup>-1</sup> (ref. 10). The GluN1b ATD and GluN1b–GluN2B ATDs were crystallized by hanging-drop vapour diffusion at 17 °C by mixing the protein (8 mg ml<sup>-1</sup>) at a 1:1 ratio with a reservoir solution containing 20% PEG3350, 150 mM KNO<sub>3</sub> and 50 mM HEPES-NaOH (pH 7.0) for GluN1b ATD, or at a 2:1 ratio with a solution containing 3.0–3.5 M sodium formate and 0.1 M HEPES (pH 7.5) for the GluN1b/GluN2B ATDs.

**Data collection and structural analysis.** Crystals were cryoprotected in buffers containing 20% PEG3350, 150 mM KNO<sub>3</sub>, 50 mM HEPES-NaOH (pH 7.0) and 20% glycerol for GluN1b ATD, or 5 M sodium formate and 0.1 M HEPES-NaOH (pH 7.5) for GluN1b–GluN2B ATDs. X-ray diffraction data were collected at the X25 and X29 beamlines at the National Synchrotron Light Source and processed using HKL2000 (ref. 23). Single anomalous diffraction data for the Se-Met-incorporated GluN1b ATD crystals were collected at the peak wavelength (0.9788 Å) and used for phasing by the program SHARP<sup>24</sup>. The initial model was built using flex-wARP<sup>25</sup>. The crystal structure of GluN1b–GluN2B ATD was solved by molecular replacement using the coordinates of GluN1b ATD and GluN2B ATD<sup>10</sup> (PDB code: 3JPW) with the program PHASER<sup>26</sup>. The models were built using COOT<sup>27</sup> and structural refinement was performed using the program PHENIX<sup>21</sup>.

**Isothermal titration calorimetry.** Proteins were dialysed overnight before the experiment against a buffer containing 150 mM NaCl, 20 mM Tris-HCl (pH 7.4) and 10% glycerol. Isothermal titration calorimetry measurements were performed using VP-ITC (MicroCal) by successive injections at 27 °C of 5  $\mu$ l of 0.15 mM ifenprodil to 0.01 mM GluN1b ATD, 10  $\mu$ l of 0.25 mM ifenprodil to 0.007 mM GluN2B ATD, 5  $\mu$ l of 0.15 mM ifenprodil to 0.01 mM GluN1b–GluN2B ATD complex and 5  $\mu$ l of 0.05 mM Ro 25-6981 to 0.007 mM GluN1b–GluN2B ATD complex. Data analysis was done using the software Origin 7.0 (Origin Labs).

**Analytical ultracentrifugation.** Sedimentation velocity and equilibrium experiments were performed using a Beckman Coulter Optima XL-I analytical ultracentrifuge. Proteins were dialysed against a buffer containing 150 mM NaCl and 20 mM Tris (pH 7.4), with or without 10  $\mu$ M ifenprodil. Sedimentation velocity experiments were performed by centrifuging protein samples loaded on 2-sector centrepieces at 42,000 r.p.m. at 20 °C. Concentration gradients were measured using interference optics or absorbance optics at a wavelength of 280 nm or 230 nm depending on the protein concentrations loaded (0.01, 0.05, 0.1, 0.5 and 1.2 mg ml<sup>-1</sup> for GluN1b–GluN2B ATD in the presence and absence of ifenprodil; 0.1 and 1.2 mg ml<sup>-1</sup> for GluN1b ATD and 0.1, 0.5 and 5 mg ml<sup>-1</sup> for GluN2B ATD). Data were analysed using the continuous c(s) and c(M) distribution models

implemented in Sedfit<sup>28</sup>. The weighted-average sedimentation coefficient ( $S_w$ ) was determined from the peak integration of c(s).

Sedimentation equilibrium experiments were performed using a 6-channel centrepiece loaded with 100- $\mu$ l protein samples at protein concentrations of 0.05, 0.1 and 0.3 mg ml<sup>-1</sup> in the presence or absence of 10  $\mu$ M ifenprodil. The samples were centrifuged sequentially at 9,000, 13,000 and 18,000 r.p.m. and allowed to reach equilibrium at each speed. Absorbance measurements were performed at wavelengths 230, 250 and 280 nm to obtain measurements at low and high protein concentrations. Global analysis of the data for multiple protein concentrations and rotor speeds was performed using single-species and A + B  $\leftrightarrow$  AB models implemented in Heteroanalysis v1.1.44 (University of Connecticut).

**Electrophysiology.** Recombinant GluN1/GluN2B NMDA receptors were expressed by co-injecting 0.1–0.5 ng of wild-type or mutant rat GluN1 and GluN2B cRNAs into defolliculated *Xenopus laevis* oocytes. The two-electrode voltage-clamp recordings were performed using agarose-tipped microelectrodes (0.4–1.0 M $\Omega$ ) filled with 3 M KCl at a holding potential of –40 mV. The bath solution contained 5 mM HEPES, 100 mM NaCl, 0.3 mM BaCl<sub>2</sub> and 10 mM Tricine at pH 7.4 (adjusted with KOH). Currents were evoked by the application of glycine and L-glutamate at 100  $\mu$ M each. Inhibition by ifenprodil was monitored in the presence of agonists and various concentrations of ifenprodil. For redox experiments, the oocytes were preincubated in the bath solution supplemented with 2 mM DTT for 3 min before recording in the continuous presence of 2 mM DTT. Data were acquired and analysed by the program Pulse (HEKA).

**Cysteine crosslinking and western blot.** Single point mutations were incorporated into the genes encoding full-length rat GluN1-4b and GluN2B in the pCI vector (Promega). Human embryonic kidney 293 cells were transfected by Eugene HD (Roche) with a mixture of 0.5  $\mu$ g of the GluN1-4b plasmid and 1  $\mu$ g of the GluN2B plasmid. Cells were harvested 24–48 h after transfection and resuspended in a buffer containing 20 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1% dodecyl-maltoside and a protease-inhibitor cocktail (Roche), as previously described<sup>29</sup>. After centrifugation at 150,000g, the supernatant was subjected to SDS–polyacrylamide gel electrophoresis (4–15%) in the presence or absence of 100 mM DTT. The proteins were transferred to Hybond-ECL nitrocellulose membranes (GE Healthcare). The membranes were blocked with TBST (20 mM Tris-HCl (pH 7.4), 150 mM NaCl and 0.1% Tween-20) containing 10% milk, then incubated with mouse monoclonal antibodies against GluN1 (MAB 1586, Millipore) or GluN2B (Invitrogen), followed by HRP-conjugated anti-mouse antibodies (GE Healthcare). Protein bands were detected by ECL detection kit (GE Healthcare).

22. Fitzgerald, D. J. *et al.* Protein complex expression by using multigene baculoviral vectors. *Nature Methods* **3**, 1021–1032 (2006).
23. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
24. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
25. Cohen, S. X. *et al.* Towards complete validated models in the next generation of ARP/wARP. *Acta Crystallogr. D* **60**, 2222–2229 (2004).
26. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
27. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
28. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
29. Furukawa, H., Singh, S. K., Mancusso, R. & Gouaux, E. Subunit arrangement and function in NMDA receptors. *Nature* **438**, 185–192 (2005).

## CORRIGENDUM

doi:10.1038/nature10220

## Thermal history of Mars inferred from orbital geochemistry of volcanic provinces

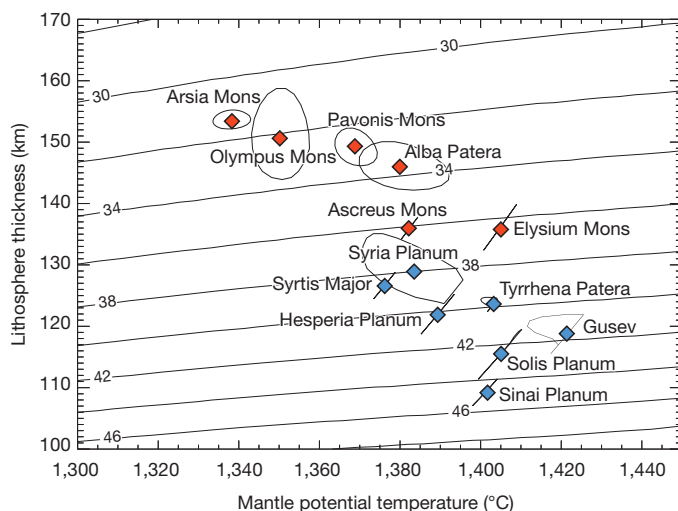
David Baratoux, Michael J. Toplis, Marc Monnereau & Olivier Gasnault

*Nature* **472**, 338–341 (2011)

An error in the calculation of heat flow contours drawn in Fig. 4 of our Letter was drawn to our attention by J. Ruiz (Universidad Complutense de Madrid). A surface temperature of 200 °C was mistakenly used instead of the correct temperature of 220 K. With the stated value of thermal conductivity ( $3.5 \text{ W m}^{-1} \text{ K}^{-1}$ ), heat flows are about 20% higher than those originally shown. The knock-on effect of this fact is that the calculated Urey ratio is 20% lower than stated, but still comfortably above the terrestrial value. On the other hand, we note that the value of thermal conductivity relevant to planetary mantles is not well constrained, with preferred values covering the range 2.5 to  $4 \text{ W m}^{-1} \text{ K}^{-1}$  (ref. 1). A value of  $3.0 \text{ W m}^{-1} \text{ K}^{-1}$  and the correct surface temperature leads to calculated heat flow very similar to those shown in our original Fig. 4. A corrected version of Fig. 4, using a surface temperature of 220 K and a conductivity of  $3.5 \text{ W m}^{-1} \text{ K}^{-1}$ , is shown below.

Figure 4 has also been corrected in the original HTML and PDF.

1. Breuer, D. & Moore, W. B. in *Treatise of Geophysics* Vol. 10, 299–341 (Elsevier, 2011).





## CORRIGENDUM

doi:10.1038/nature10223

### Telomere dysfunction induces metabolic and mitochondrial compromise

Ergün Sahin, Simona Colla, Marc Liesa, Javid Moslehi, Florian L. Müller, Mira Guo, Marcus Cooper, Darrell Kotton, Attila J. Fabian, Carl Walkey, Richard S. Maser, Giovanni Tonon, Friedrich Foerster, Robert Xiong, Y. Alan Wang, Sachet A. Shukla, Mariela Jaskelioff, Eric S. Martin, Timothy P. Heffernan, Alexei Protopopov, Elena Ivanova, John E. Mahoney, Maria Kost-Alimova, Samuel R. Perry, Roderick Bronson, Ronglih Liao, Richard Mulligan, Orian S. Shirihai, Lynda Chin & Ronald A. DePinho

*Nature* **470**, 359–365 (2011)

In this Article, references<sup>1–3</sup> reporting an association between telomere dysfunction and mitochondrial impairment or TERT deficiency and mitochondrial impairment were inadvertently omitted.

1. Passos, J. F. *et al.* Feedback between p21 and reactive oxygen production is necessary for cell senescence. *Mol. Syst. Biol.* **6**, 347, doi:10.1038/msb.2010.5 (2010).
2. Haendeler, J. *et al.* Mitochondrial telomerase reverse transcriptase binds to and protects mitochondrial DNA and function from damage. *Arterioscler. Thromb. Vasc. Biol.* **29**, 929–935 (2009).
3. Kovalenko, O. A. *et al.* A mutant telomerase defective in nuclear-cytoplasmic shuttling fails to immortalize cells and is associated with mitochondrial dysfunction. *Aging Cell* **9**, 203–219 (2010).

CORRIGENDUM

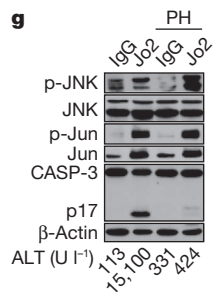
doi:10.1038/nature10221

## CD95 promotes tumour growth

Lina Chen, Sun-Mi Park, Alexei V. Tumanov, Annika Hau, Kenjiro Sawada, Christine Feig, Jerrold R. Turner, Yang-Xin Fu, Iris L. Romero, Ernst Lengyel & Marcus E. Peter

*Nature* **465**, 492–496 (2010)

In our recent Corrigendum (*Nature* **471**, 254 (2011); doi:10.1038/nature09897), Fig. 4g inadvertently contained three incorrect panels. The corrected Fig. 4g is shown below. This mistake does not alter the overall conclusions of this Letter.



# CAREERS

**COLUMN** Administrative tasks keep scientists from doing science **p.257**

**ONLINE JOURNAL** Job markets are tight, but PhDs bring potential [go.nature.com/3fhtcj](http://go.nature.com/3fhtcj)

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)



COMMUNICATION

## The best words in the best order

*Those who prefer organizing ideas to working at the bench should consider a career in technical writing.*

BY LAURA BONETTA

The words ‘technical writer’ evoke an image of someone sitting in a windowless office composing a manual for a DVD player. But the field affords many opportunities not nearly so austere and mundane.

These days, technical writers provide information about products and services not just

as instruction manuals, but through websites, e-learning materials, online help modules and FAQ pages, wikis, podcasts and blogs. And they focus on a range of projects — from composing step-by-step protocols for setting up an electron microscope and using the imaging software, to writing scientific manuscripts and regulatory documents. These writers work not in isolation, but as part of teams of

researchers, engineers, physicians or computer scientists.

Owing to the diversity of their duties, technical writers often refer to themselves as technical communicators; if they write about drugs or medical devices, they are medical communicators. Their tasks sometimes overlap with those of public-relations officers, authors and other types of writers, but technical communicators are set apart by their focus on presenting factual information about complex subjects as clearly and efficiently as possible.

Many technical communicators say that what they enjoy most about their job are the constant opportunities to learn — they must keep pace with new media and understand the products and services they write about. The field offers a lot of flexibility — from nine-to-five desk jobs to freelancing from home. And the job market is relatively healthy: according to the US Bureau of Labor Statistics, there were about 48,900 posts for technical writers in the United States in 2008, and the number of positions is expected to grow by more than 18% by 2018. The medical-writing market doubled between 2003 and 2008, and is also expected to continue growing. In 2009, Germany had some 80,000 people working in technical communication, according to the technical writers’ association *tekem* in Stuttgart. And the Institute of Scientific and Technical Communicators (ISTC), a professional association based in Croydon, UK, says that mobile telecommunications and pharmaceuticals are strong growth areas for technical writers in Britain.

The need to research a range of topics and grasp thorny subjects makes technical communication a great field for people with advanced scientific degrees, from newly minted PhDs to senior postdocs or even established scientists wanting to leave the lab. Of course, they must also have a talent for constructing concise, clear and organized prose.

### TECHNICAL DEFINITION

Unlike science journalism, “technical writing is not about taking something technical and making it nontechnical”, says Paul Ballard, managing director of the technical-communications firm 3di in Woking, UK. “It is about taking something that is complex and making it clear to those who need to use it.” The challenge, explains Ballard, is to understand the subject in depth and to be able to empathize with those who will be using the information — whether they are scientists, physicians, ►



► technicians or officers at regulatory agencies. “You have to be able to predict risk areas and any difficulties a user might have, and be able to answer any possible questions efficiently,” he adds.

Debbie Davy, a medical communicator in Toronto, Canada, has done everything from writing manuals for electrocardiography machines to designing software for searching through and analysing medical information gathered by hospitals. “Those reports are normally cumbersome to use and written in obscure language,” says Davy. “I provide the interface for reinterpreting the results.”

Such tasks often involve a lot of research. “The bulk of what I do is interview the client to determine what they need and then look at what they have produced in the past,” says Davy. She meets stakeholders in the company, including engineers and product developers, to gather information, then pinpoints her target audience and the best way to present the information to them. Technical communicators typically need to work with web-publishing and authoring tools, such as Adobe FrameMaker and the increasingly popular Darwin Information Typing Architecture. “A lot of technical writers get hung up on the tool of the moment,” says Davy. “You have to know [the computer coding language] XML, but anyone can learn that. My value is in the organization of ideas, the identification of the audience



***“If you want to hide in a room and not talk to anyone, it’s not a good job for you.”***

Paul Ballard

and communicating information.”

Day-to-day tasks for scientific technical communicators are typically writing descriptions of instruments, products and databases. Medical communicators write regulatory documents, protocols and procedures, summaries of efficacy and safety studies, and reports of clinical trials. Many writers also find jobs in contract-research organiza-

tions (CROs), which do research on behalf of pharmaceutical companies and academic labs. At CROs, technical communicators write up results for publication in journals, white papers and regulatory documents. Another option for medical writers is to develop and write materials for the continuing medical-education courses that physicians and other health-care professionals must take each year.

According to the Society for Technical Communication (STC), based in Fairfax, Virginia, the median salary for technical writers (both freelance and employed full time) across all

industries in the United States in 2008 was US\$61,620. A salary survey conducted in 2007 by the American Medical Writers Association (AMWA) in Rockville, Maryland, shows that median salaries for US medical writers with a master’s degree were \$77,339 for female employees and \$86,240 for male employees. Freelancers earned a median of \$85,406 for women and \$107,444 for men. For those with a PhD, employees’ figures were \$91,797 for women and \$101,872 for men; and female freelancers got \$114,692, and men got \$131,143. The ISTC says that junior technical-communicator positions in the United Kingdom have a starting salary of about £20,000 (US\$32,000) a year, and experienced practitioners earn about £40,000. Freelance rates may be up to £50 an hour.

### THE WRITE FIT?

Some technical communicators are former engineers, life scientists or computer scientists. “Many of the people we hire come in having been an engineer and having found that they like the writing aspect of their job a lot more than actually being an engineer. That applies to a lot of scientific fields,” says Ballard. About 50% of AMWA members have science or health-care degrees, including biology, pharmacy, medicine, public health and nursing; some 30% earned their highest degrees in non-science subjects, including English, journalism, communication and technical writing.

Many companies and organizations put a premium on technical communicators with advanced degrees in relevant subjects. “Subject-matter expertise and knowledge of the audience are priceless,” says Andrew Davis, a recruiter at Content Rules, a content-development company in Santa Clara, California. Davis has hired technical communicators for laboratory-equipment developers and suppliers such as Life Technologies in Carlsbad, California; Affymetrix in Santa Clara; and BD Biosciences, based in Franklin Lakes, New Jersey. “Those companies would not consider a candidate without a master’s or PhD,” says Davis. Technical communicators’ tasks at such firms range from writing online manuals for analytical tools for the chemical and pharmaceutical industries to composing instructions for using molecular-analysis databases, he says.

A science degree signals to employers that the applicant can understand complex scientific materials, think critically, gather and synthesize lots of data, and analyse and interpret information, Davis notes. But the most important skill is communicating effectively, both in writing and verbally. Most fledgling technical writers build on communication skills that they started to hone as part of their research training.

Technical communicators need to be able to work independently, but the job is not solitary. “If you want to hide in a room and not talk to anyone, it’s not a good job for you,” says Ballard. “Good technical authors have to communicate with people face to face or on Skype chats.”

## GETTING TRAINED

### Degrees in technical writing

#### United States

● **Auburn University, Alabama**  
Master of technical and professional communication  
[www.cla.auburn.edu/mtpc](http://www.cla.auburn.edu/mtpc)

● **Boise State University, Idaho**  
MA in technical communication  
[go.nature.com/ad3ubz](http://go.nature.com/ad3ubz)

● **Illinois Institute of Technology, Chicago**  
MS in technical communication and information design  
[www.iit.edu](http://www.iit.edu)

● **Mercer University, Penfield, Georgia**  
MSc in technical communication management  
[www.mercer.edu](http://www.mercer.edu)

● **Missouri University of Science and Technology, Rolla**  
MSc in technical communication  
[go.nature.com/l4oc1p](http://go.nature.com/l4oc1p)

● **Northeastern University College of Professional Studies, Boston, Massachusetts**  
MSc in technical communication  
[go.nature.com/tyb4fl](http://go.nature.com/tyb4fl)

● **University of Houston – Downtown, Texas**  
MSc in professional writing and technical communication  
[go.nature.com/xlxyve](http://go.nature.com/xlxyve)

● **University of Minnesota, Minneapolis**  
MS in scientific and technical communication  
[go.nature.com/wjry7m](http://go.nature.com/wjry7m)

● **University of Wisconsin–Stout**  
MSc in technical and professional communication  
[go.nature.com/1hs9ow](http://go.nature.com/1hs9ow)

● **University of Texas Southwestern Medical Center, Dallas**  
MA in biomedical communications and biomedical illustration  
[go.nature.com/usjj7a](http://go.nature.com/usjj7a)

**United Kingdom**  
● **Sheffield Hallam University**  
MA in technical communication  
[www.shu.ac.uk](http://www.shu.ac.uk)

● **University of Portsmouth**  
MA in technical communication  
[www.port.ac.uk](http://www.port.ac.uk)

Flexibility is also key. Most technical writers start off working in their areas of expertise but have to take on projects outside that field, depending on what is most in demand.

### BREAKING IN

Many companies are willing to hire novice technical communicators who have a PhD in the right field, but employers want to see evidence of good writing skills. "Anything you have already written can be used, such as a newsletter or a website, and it does not have to be published," says Lori Alexander, president of Editorial Rx, a medical-writing company in Orange Park, Florida, and editor of the *AMWA Journal*. Applicants can also use research abstracts or articles as writing samples.

Societies including the STC, AMWA and ISTC organize conferences and workshops for people interested in technical communication. "The most important thing is to network," says Alexander. It might pay to take some courses in technical or medical communication. "These help demonstrate to an employer that you are committed to becoming a writer," says Ballard.

Julie Gelderloos earned a PhD in cellular and developmental biology before becoming a medical writer based in Boulder, Colorado. While working, she earned AMWA certification by attending courses on writing skills, ethics, statistics, pharmacology, epidemiology and the business aspects of freelancing. "I still regularly attend AMWA conferences and learning sessions to keep up with changes in the field and to network," she says. Although these courses helped Gelderloos to hone her skills, she says that most of her training was acquired on the job.

For those seeking more in-depth training, several universities offer certificates and master's degrees in technical communication (see 'Degrees in technical writing'). Such qualifications are "definitely a bonus when applying for a job," says Gelderloos.

But ultimately, Gelderloos says, most employers look for writing experience first. And when hiring junior writers herself, Gelderloos made a point of finding candidates with some experience in medical writing.

Technical-communication programmes can, however, serve as a pathway to internships at companies, which can be invaluable for gaining experience in the field. They also provide a quick way to find out whether the profession is a good fit for the aspirant.

Experience in the lab will provide clues to this. "If you find you like doing the writing rather than the science aspect, you like interpreting the data rather than generating them, and you are the person everyone goes to because they need something written," says Alexander, "this might be a good career for you." ■

**Laura Bonetta** is a freelance writer based in Garrett Park, Maryland.

## COLUMN

# Too many tasks

*A heavy administrative burden keeps top academic scientists from doing science, argues Adam James.*

When I grow up, I want to be a scientist. Sometimes, though, I fear that it is no longer possible. I fear that I will become an administrator of scientific tasks rather than an investigator of scientific truths.

My academic experience thus far has taught me that science has become something done mainly by graduate students and postdoctoral researchers, not by the principal investigators who have ostensibly achieved their career goal — the freedom to steer and delve into their own research ideas. The students and postdocs might do science with the firm, guiding hand and skilful collaboration of the principal investigator, or in a series of inspired individual moments, deliberate or otherwise. But they are the ones most often practising the craft. The principal investigator isn't a scientist — not anymore. This person, with all of the education and training to be a scientist, has instead become an administrator.

The tasks are many. Principal investigators must apply for funding, a regular and obviously essential, yet time-consuming, exercise. They must stock the lab with equipment, nurture ideas about the research group's direction and, optimally, relay knowledge about how to use equipment and manage data. They're also often lecturers, or unit or department coordinators, and perhaps advisers for undergraduate students. The principal investigator, as I see it from my graduate-student perch, also provides the link between the lab and the research world by supplying junior scientists with a network of scientific colleagues, financial support to attend conferences, guidance on writing papers and other key forms of help.

In the middle of (and despite) all these duties, he or she should also, ideally, be a source of inspiration for the next generation of scientists. The principal investigator should spend some time supervising undergraduate laboratories; after all, when better to establish a rapport with potential researchers than during long hours assisting with their studies?

The demands are many, and most are worthy. They could easily fill the time of anyone who cares enough to pursue them to a satisfactory, let alone rigorous, completion. Why are course coordination and other routine administrative tasks part of the job description of tenured academics? If our top researchers weren't burdened with these duties, perhaps



IMAGES.COM/CORBIS

they could get back into the lab and make more of their contributions to science.

My view of science might be idealistic. To me, science is the way to directly improve the progress and future of humanity. Many occupations contribute to or enhance our lives, from manufacturing and production to sports and entertainment. Politics is admirable in many respects as an archetype of selfless giving, although it is tainted by individual failings and strategies aimed strictly at winning elections. But it is science that provides the means to make a long-term, meaningful contribution to our future by shining a light down the often-dark halls of knowledge. If the scientist inevitably becomes an administrator, this lofty goal seems much harder to achieve.

Somewhere along the line, the institution of scientific research began overloading scientists with too many tasks, too many responsibilities. Hence, the graduate students who do become inspired with the wonder of science may go on to become administrators — less connected to the very science that inspired them. We are training to become scientists, yet, on successfully reaching our goal, we are promptly promoted out of our primary objective and passion.

Maybe this constellation of administrative and research tasks is simply the way things are in an increasingly complex and bureaucratic scientific research environment. Nevertheless, it is a concern for the budding scientist. We need more scientists. We need more teachers and administrators. Ideally, they would not all be the same person. ■

**Adam James** is a PhD candidate in synthetic chemistry at the University of Tasmania in Hobart, Australia.



# THUMBS

*It's all relative.*

BY GORDON CASH

It started as party chatter.

My hands are different sizes, not so much that most people can tell, but I can. The real difference is the thumbs. The distal joint of my right thumb, but only the right one, bends back 90°, 'hitchhiker's thumb'. The medical term is distal hyperextensibility. It's an autosomal recessive trait, so theoretically you can't have just one. At parties, I said, "I have someone else's hand. When they were putting me together, somebody dropped a hand and stepped on it, so they had to go for spare parts." Great conversation starter, for a while.

Then I read about the Chimaera Project.

A chimaera is an organism composed of two genetically distinct cell lines. Natural chimaeras occur when fraternal twin embryos fuse early enough to develop into a single entity, with no extra limbs or organs.

I saw a TV show once where a woman sought legal custody of two children on the grounds that she was their biological mother. The problem was, DNA tests kept showing that she wasn't. Finally, her doctor figured out she was a chimaera. One of her genomes had been tested; the other one was the children's mother. The doctor told the woman's lawyer this was a rare condition, but he asked how she knew that, as no one ever tested for it. I was living in an urban area with several medical research institutions. Anonymous donors funded one of them to find out how common chimaeras really were. Maybe the donors saw the same TV show.

The clinic where the Chimaera Project was based was nearby, and I managed to get an appointment. Dr Richard Mott was intrigued enough by my thumbs to take some samples from them. He said my results would be ready in a few weeks, but three days later, his assistant called and asked me to come in. When Dr Mott asked to see me personally, I thought he just wanted more samples. Instead, he had my results.

"You were an easy case, Mr Treadwell. Your left hand is female."

When I was bemused instead of horrified, he told me that opposite-gender chimaerism goes unnoticed unless genitals or hormone-secreting organs are involved. When he remarked on how calmly I took the news, I assumed he had other patients who took it

badly. He still wanted me to see a counsellor.

Dr Mary Austin was pleasant enough, and she really did seem concerned about me. Towards the end of the session, she asked if I felt like two people.

"I'm the same person I've always been, Dr Austin. I never bought the unique genome story, even when the anticyclones came."

Anticyclone is a meteorological term,



co-opted to mean someone who opposed cytogenetic cloning — anti-cy-clone. A lame coinage, certainly, but the bloggers immortalized it. I could tell she hadn't expected me to know about that, but not whether she was pleased or upset.

"Two people? Does anyone actually believe that?"

"You'd be surprised. You actually remember the anticyclones?"

"Barely. I was a child, but I remember the mobs, and the bodies in the streets. And the older kids told stories." The older kids told of surgical scars ripped open to recover cloned

organs, not because they were worth any money, but to give them decent burials. I thought they were just making it up to

scare us, as older kids do, but I have since read the histories. Know-nothings insisted for decades that a zygote was a person because it had a unique genome; then they discovered identical twins. Thus was the anticyclone movement born. Because the cloned cell could theoretically grow into the patient's twin instead of an organ, they insisted the organ was a separate person, too.

"Why would I be surprised?"

"Maybe you wouldn't be, Mr Treadwell, as you remember the anticyclones," was all she said. At least I had some better party chatter to go with my thumbs.

The Chimaera Project must have been a scientific success because before too many more parties, I found out it had generated a backlash. It seemed there were quite a few of us, and we had attracted the wrong sort of attention. Some people thought we were abominations.

I liked to think the anticyclone movement withered because the public finally saw the value of all those people who lived because they could regenerate a damaged organ, rather than died. Cynics said the movement perished of intellectual bankruptcy. It didn't matter. Like the mythical hydra that grew two heads where one had been lopped off, this beast would not be slain so easily.

The evening news said Dr Mott's clinic was burning. Dr Austin's husband told the police she never came home from work yesterday. Her office had been ransacked, and her patient files were missing. The brick that came through my bedroom window later that week was wrapped in a piece of paper that said: "And if thy hand offend thee, cut it off: it is better for thee to enter into life maimed, than having two hands to go into hell, into the fire that never shall be quenched."

The police treated my incident as the work of a lone vandal or maybe a prank, certainly not the tip of an iceberg decades deep. The older kids of my childhood told another scary truth: there is no place to hide from the madness. So, I shall fight it where I can and crush it where it rises. My one-time sister's hand will strike the blows. The party is over. ■

**Gordon Cash** works as a chemist in Washington DC. He has no idea whether his mismatched thumbs are due to chimaerism, or to some other anomaly of development.

JACEY

➔ **NATURE.COM**  
Follow Futures on  
Facebook at:  
[go.nature.com/mtoodm](https://www.facebook.com/mtoodm)



# Competition, predation and natural selection in island lizards

ARISING FROM R. Calsbeek & R. M. Cox *Nature* **465**, 613–616 (2010)

Discerning the relative influence of competition and predation as selective forces is an important goal of evolutionary ecology. Calsbeek and Cox<sup>1</sup> argue that intraspecific competition outweighs predation as an agent of natural selection on island populations of the lizard *Anolis sagrei*. However, we identify several problems with the design and analysis of the Calsbeek and Cox<sup>1</sup> study that we believe render its results uninterpretable.

Calsbeek and Cox<sup>1</sup> manipulated lizard population density and predator occurrence on four small islands. The predation manipulation had three treatments: ‘none’ (netting covering islands to exclude birds); ‘birds’ (netting placed around the perimeter of, but not covering, islands, allowing bird access); and ‘birds and snakes’ (three snakes added to islands without any netting). Lizards were introduced onto islands such that each predation treatment was paired once with a ‘high’ and once with a ‘low’ density treatment, although statistical analyses treated density as a continuous variable (contrary to the impression given by their Fig. 2). Over two years (2008 and 2009), each possible combination of the three predation treatments and two lizard-density treatments was established one time (each trial lasting 4 months, with two islands used in both years and two in 2009 only). In addition, one unmanipulated island, Kidd Cay (Fig. 1a), was included in 2008. On each island, the authors recorded survival, habitat use and natural selection on several traits.

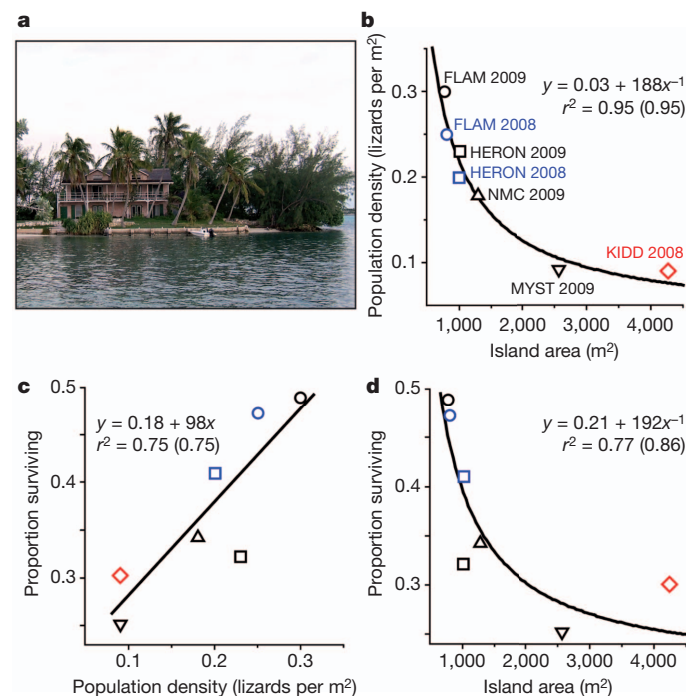
This experimental design is confounded in three fundamental ways. First, density is confounded with island area. All analyses treat lizard density as a surrogate for intraspecific competition. However, an inverse correlation with island area explains 95% of the variation in density (Fig. 1b), such that it is impossible to disentangle the two factors statistically. This is a crucial problem, because multiple factors related to both predation and competition are known to vary with island area. For example, as island area increases, so too do the number of bird species<sup>2,3</sup> (which increases the number of potential predators) and mean vegetation height<sup>3</sup> (which might increase lizards’ susceptibility to avian predation<sup>4</sup>). Likewise, because larger islands have lower perimeter/area ratios, they receive relatively lower input of marine-resource subsidies and have lower arthropod densities<sup>5</sup>; a study of *A. sagrei* in this system showed that lizard densities vary significantly with the amount of seaweed deposition, and that experimental seaweed deposition increased lizard densities by more than 60% (ref. 6).

Because of these relationships, there is no way to distinguish the relative importance of ‘competition’ (that is, density) versus predation in driving the results. This point is illustrated by the fact that density and island area are essentially equivalent predictors of both survival (Fig. 1c, d) and year-corrected selection differentials ( $r^2 = 0.66$  versus 0.74, respectively, for snout–vent length; 0.50 versus 0.52 for hindlimb length; and 0.92 versus 0.83 for stamina;  $r^2$  values drawn from the better-fitting regression of  $y$  against either  $x$  or  $x^{-1}$ ). Thus, we believe that the density–area correlation alone invalidates the conclusion<sup>1</sup> that intraspecific competition drives selection on *A. sagrei*.

There is another problem with the claim<sup>1</sup> that intraspecific competition caused the observed variability in selection differentials. If competition for resources drives natural selection, then greater densities should have negative effects, such as lower survival, on individuals. However, our re-analysis of the data shows that survival is actually

positively correlated with density (Fig. 1c). Thus, the assumption that density is a proxy for competition—which underpins the entire study<sup>1</sup>—seems unwarranted. Indeed, we can think of no plausible causal explanation for a direct positive relationship between density and survival in *A. sagrei*. Instead, we suspect that this relationship is an artefact of the near-perfect correlation between density and island area (Fig. 1b), and that differences in survival are actually driven by indirect island–area effects (Fig. 1d). As discussed above, previous work suggests multiple explanations for an inverse relationship between lizard survival and island area<sup>2–6</sup>, any or all of which might have operated in this study.

The second structural flaw in the design is the confounding of treatment with year. The birds-and-snakes treatment was applied only in 2009, which makes it impossible to separate the effects of snake addition from the effects of year. The authors controlled for year effects in their analyses of selection by analysing residuals of the regression of selection differentials against year. But because snake addition was only conducted in one year, removing year effects also partially removes any effects of snake addition.



**Figure 1 | Confounding relationships among treatments and variables.** **a**, Kidd Cay, which differs markedly from other islands used in the study (compare with Supplementary Fig. 1a of Calsbeek and Cox<sup>1</sup>). **b**, Strong inverse correlation between lizard density and island area. Islands re-used in successive years share the same symbol (names from Calsbeek and Cox<sup>1</sup>), with those from 2008 shown in blue and Kidd Cay shown in red;  $r^2$  value in brackets is that obtained when Kidd Cay is excluded from the analysis. **c**, Positive correlation between density and survival, probably an artefact stemming from the inverse relationships between density and area (**b**) and survival and area (**d**). All analyses use data presented in or calculated from Supplementary Table 1 of Calsbeek and Cox<sup>1</sup>.

Moreover, the correction for year was applied inconsistently. Ordinarily, such a correction would be applied only to variables showing significant variability between years. Instead, Calsbeek and Cox<sup>1</sup> removed year effects in all analyses of selection differentials, even though only one variable differed significantly from 2008 to 2009 (stamina: ANOVA,  $F_{1,5} = 26.2$ ,  $P < 0.004$ ). In contrast, year effects were not removed from analyses of survival, even though male survival was 50% lower in 2009 than in 2008 ( $F_{1,4} = 24.3$ ,  $P = 0.007$ ; also see below). Calsbeek and Cox<sup>1</sup> report that the predation treatments reduced survival, but if year effects had been removed from the survival analyses, they would have found no significant effect of predation on survival. Conversely, the authors report significant effects of population density on selection differentials after correcting for year effects, but none of the selection differentials is significantly related to density when year effects are not removed (generalized linear models with normal distribution and identity link function using JMP 8.02 software, all  $P > 0.05$ ). Thus, the inconsistent handling of year effects determines the main conclusions of the paper, whereas a consistent approach to year effects would have failed to provide support for one or the other set of conclusions.

The third confounding relationship in the design of the Calsbeek and Cox study<sup>1</sup> is between year and sex ratio. The two islands manipulated in 2008 were seeded with 40 males and ~160 females (1:4 sex ratio), whereas the four islands manipulated in 2009 received ~80 males and ~150 females (1:1.9 ratio). Male *A. sagrei* are very aggressive towards conspecific males; therefore, greater male/female ratios might lead to increased agonistic behaviour between males, which is energetically costly and likely to increase predation risk<sup>7</sup>. Thus, the 100% increase in male/female ratio in 2009 might have caused the aforementioned 50% decrease in male survival observed in that year (regression of survival against sex ratio:  $r = -0.93$ ,  $F_{1,4} = 25.6$ ,  $P = 0.008$ ).

Calsbeek and Cox attributed survival differences to predation because survival was lowest on islands with birds and snakes (Fig. 1 of ref. 1). What we show above is that because the two islands with both birds and snakes existed only in 2009, and therefore received twice as many males per female as islands manipulated in 2008, it is impossible to make any causal inference about variation in male survival. The observed variation might have been caused by effects of sex ratio, by environmental differences across the two years of the study, by the predation manipulation, or by some combination or interaction of these three factors.

In addition to these three confounding relationships, a fourth problem in the study design involves the unmanipulated island Kidd Cay, which was “monitored...as a natural reference population”. However, Kidd Cay was used as more than just a reference point, because it was included in statistical analyses of selection strength as a ‘bird-only’ island. (However, it was not included in analyses of lizard survival, which would have eliminated the reported effect of predation on survival<sup>1</sup>.) The inclusion of this unmanipulated island in tests of the experimental effects of density and predation is inappropriate because Kidd Cay is qualitatively different from the experimental islands (Fig. 1a): it is much larger, has a hotel on it, is connected by a causeway to an even larger island, and supports domestic predators, *Anolis* species other than *A. sagrei*, and large trees (some  $\geq 10$ -m tall), none of which occurs on the experimental islands (where most trees are  $< 3$ -m tall).

Additional concerns include the absence of information necessary to replicate several of the analyses (for example, the final two sentences of the Methods describe analyses not reported in the paper, and no

substantive methods are provided for the analyses in Table 1); statistical non-independence of replicates resulting from the re-use of two islands in successive years; potential biases arising from the use of  $AIC_c$  to compare models with 3–5 parameters when  $n = 7$ ; and failure to control for the effect of placing netting on islands in the birds-and-snakes treatment, which confounds the presence of snakes with the absence of netting. Any of these issues might have influenced the results. However, the web of confounding correlations among the variables (especially the statistical near-equivalence of density and island area and the positive density–survival relationship, which together invalidate the use of density as a proxy for competition) means that neither post-hoc statistical palliatives nor the exclusion of Kidd Cay from analyses can resolve the relative importance of competition and predation as agents of selection in this experiment.

The recent advent of experimental field studies in evolution promises investigation of theoretical predictions once thought untestable. In conducting such field studies, however, evolutionary biologists must ensure adequate replication, include appropriate controls for all manipulations, and scrutinize potentially confounding correlations between variables. Ecologists have grappled with these issues for decades, and the ecological literature offers guidance for dealing with them. We sympathize with the difficulties of conducting large-scale field experiments, and we applaud both the vision that motivated this study and the inclusion of the raw data in the Supplementary Information, but unfortunately those data cannot answer the central question posed by Calsbeek and Cox<sup>1</sup>. Thirty years of research in this Bahamian island system suggest that both competition and predation can influence selection, yet we still await a robust experimental test of their relative importance.

**Jonathan B. Losos<sup>1</sup> & Robert M. Pringle<sup>2,3</sup>**

<sup>1</sup>Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA.  
e-mail: jlosos@oeb.harvard.edu

<sup>2</sup>Society of Fellows, Harvard University, Cambridge, Massachusetts 02138, USA.

<sup>3</sup>Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA.

**Received 20 September 2010; accepted 4 April 2011.**

1. Calsbeek, R. & Cox, R. M. Experimentally assessing the relative importance of predation and competition as agents of selection. *Nature* **465**, 613–616 (2010).
2. Schoener, T. W. & Schoener, A. Distribution of vertebrates on some very small islands. II. Patterns in species number. *J. Anim. Ecol.* **52**, 237–262 (1983).
3. Schoener, T. W. & Schoener, A. Inverse relation of survival of lizards with island size and avifaunal richness. *Nature* **274**, 685–687 (1978).
4. Schoener, T. W., Losos, J. B. & Spiller, D. Island biogeography of populations: an introduced species transforms survival patterns. *Science* **310**, 1807–1809 (2005).
5. Polis, G. A. & Hurd, S. D. Linking marine and terrestrial food webs: allochthonous input from the ocean supports high primary productivity on small islands and coastal land communities. *Am. Nat.* **147**, 396–423 (1996).
6. Spiller, D. A. *et al.* Marine subsidies have multiple effects on coastal food webs. *Ecology* **91**, 1424–1434 (2010).
7. Díaz-Uriarte, R. Anti-predator behaviour changes following an aggressive encounter in the lizard *Tropidurus hispidus*. *Proc. R. Soc. Lond. B* **266**, 2457–2464 (1999).

**Author Contributions** J.B.L. and R.M.P. analysed data and wrote the paper.

**Competing financial interests:** declared none

doi:10.1038/nature10140

## Calsbeek &amp; Cox reply

REPLYING TO J. B. Losos & R. M. Pringle *Nature* **475**, doi:10.1038/nature10140 (2011)

We agree with several of the points raised by Losos and Pringle<sup>1</sup>, but we show here that our data<sup>2</sup> still implicate competition as an agent of natural selection, while providing only limited support for a role of predation. Although patterns of density-dependent survival and selection on Kidd Cay are highly congruent with those on experimental islands, this site could be considered fundamentally different. We therefore base our rebuttal on analyses that use mean values per island (as in our original paper<sup>2</sup>) but which now exclude Kidd Cay ( $n = 6$  experimental islands).

We agree that an ideal experimental design would balance predator treatments across years, but we note that the benefits of large-scale experiments often outweigh necessary sacrifices in replication<sup>3,4</sup>. Given that survival was higher in 2008 than in 2009, some of the treatment effects on survival in our Fig. 1 (ref. 2) do reflect year effects. However, during 2009, the year in which all predator treatments were included, survival of males still tended to be lower on islands exposed to bird and snake predators (mean survival = 0.20, 0.30) than on other islands (0.34, 0.35) (generalized linear model (GLM) with identity link function:  $\chi^2 = 2.68$ ,  $P = 0.10$ ;  $n = 4$ ). More importantly, there is no evidence that predators influenced selection on any trait, whether year effects are included (GLM: all  $P > 0.26$ ) or excluded (GLM: all  $P > 0.33$ ). The same is true when analyses are restricted to 2009 (GLM: all  $P > 0.21$ ;  $n = 4$ ). When individual survival (0 or 1) is analysed as the response variable, predator treatment affects overall survival in 2009 (GLM with logit link:  $\chi^2 = 47.59$ ,  $P < 0.0001$ ;  $n = 323$ ), yet no treatment  $\times$  phenotype interactions are significant (GLM: all  $P > 0.43$ ). This comparison does not provide replication at the population level, but it strongly suggests that predators had little effect on phenotypic selection.

By contrast, population density tends to be associated with phenotypic selection regardless of whether density is treated as a categorical (high/low) variable (GLM all with identity link for snout–vent length:  $\chi^2 = 3.15$ ,  $P = 0.08$ ; hindlimb length:  $\chi^2 = 3.32$ ,  $P = 0.07$ ; stamina:  $\chi^2 = 5.01$ ,  $P = 0.03$ ) or as a continuous variable in analyses including year effects (GLM for snout–vent length:  $P = 0.25$ ; hindlimb length:  $P = 0.04$ ; stamina:  $P < 0.001$ ). Moreover, a two-factor GLM (identity link) with predator and density treatments reveals a significant effect

of density, but not predators, for selection on snout–vent length (density:  $\chi^2 = 4.51$ ,  $P = 0.03$ ; predators:  $\chi^2 = 2.46$ ,  $P = 0.29$ ) and stamina (density:  $\chi^2 = 6.02$ ,  $P = 0.01$ ; predators:  $\chi^2 = 4.03$ ,  $P = 0.13$ ). Both density (GLM  $\chi^2 = 17.17$ ,  $P < 0.001$ ) and predators (GLM  $\chi^2 = 15.42$ ,  $P < 0.001$ ) influenced selection on hindlimb length, but predator effects occur because selection was only observed in the absence of predators. Therefore, analyses that exclude Kidd Cay and use only uncorrected selection differentials support the main conclusion of our paper<sup>2</sup> by showing that density influenced selection on each of these traits, whereas predators had little effect on selection.

A more general issue is the extent to which lizard density can be interpreted as a surrogate for competition. We agree that the positive correlation between survival and density (Losos and Pringle<sup>1</sup>, Fig. 1c) challenges this assumption. However, the focal result of our study was to show that the relationship between survival and phenotype changed as a function of density, not that overall mortality differed among treatments. Losos and Pringle<sup>1</sup> show that island area is correlated with density and propose the interesting alternative hypothesis that selection could be driven by factors related to island area. We see no reason to consider this a more parsimonious interpretation at present, but we agree that future experiments must explicitly disentangle the effects of density and island area.

**Ryan Calsbeek<sup>1</sup> & Robert M. Cox<sup>1</sup>**

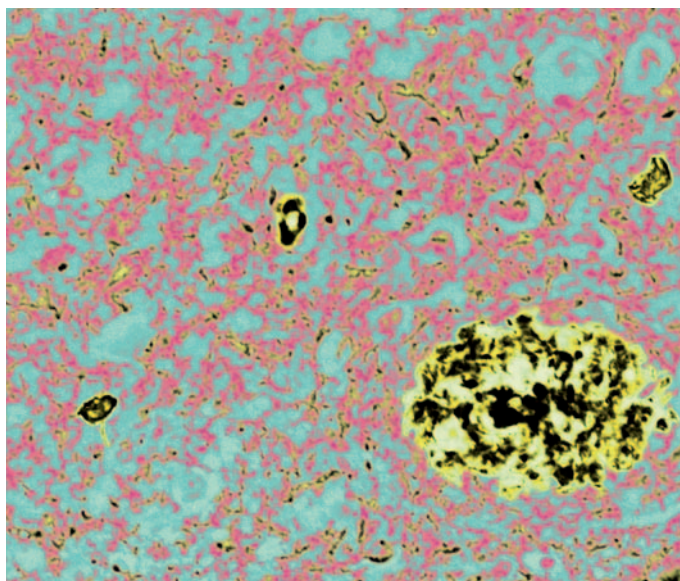
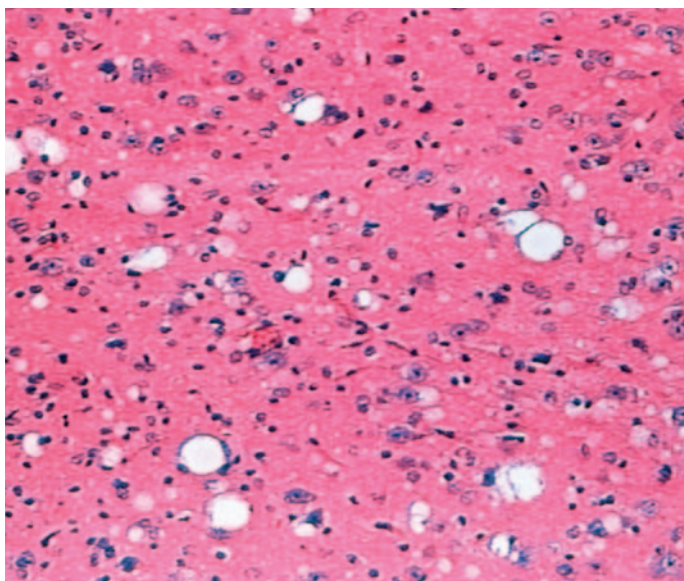
<sup>1</sup>Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire 03755, USA.

e-mail: ryan.calsbeek@dartmouth.edu

1. Losos, J. B. & Pringle, R. M. Competition, predation and natural selection in island lizards. *Nature* **475**, doi:10.1038/nature10140 (2011).
2. Calsbeek, R. & Cox, R. M. Experimentally assessing the relative importance of predation and competition as agents of selection. *Nature* **465**, 613–616 (2010).
3. Oksanen, L. Logic of experiments in ecology: is pseudoreplication a pseudoissue? *Oikos* **94**, 27–38 (2001).
4. Carpenter, S. R. Large-scale perturbations: Opportunities for innovation. *Ecology* **71**, 2038–2043 (1990).

doi:10.1038/nature10141





SIMON FRASER/JAMES KING-HOLMES/SCIENCE PHOTO LIBRARY

At autopsy, the brains of Alzheimer's patients (right) are filled with amyloid plaques, reminiscent of the plaques seen in the brains of animals with scrapie (left).

## AMYLOID

# Little proteins, big clues

*After a quarter of a century, the amyloid hypothesis for Alzheimer's disease is reconnecting to its roots in prion research.*

BY JIM SCHNABEL

In September 1984, a group of prominent researchers from around the world met in Scotland to discuss a disease that afflicted sheep and goats.

Scrapie, as they called it, was important for more than agricultural reasons — it was also the most easily studied example of an emerging class of diseases that destroyed the brain. The illnesses jumped infectiously from animal to animal, yet yielded no trace of a virus or other microorganism. One big clue was that these diseases left behind insoluble clumps, or plaques, made from millions of tiny fibrils, each of which comprised hundreds or thousands of proteins. A striking new hypothesis proposed that these fibrils and their plaques marked the toxic passage of infectious proteinaceous particles, or prions.

On the first night of the conference, several researchers gathered for dinner. Among them were Colin Masters, a neuropathologist from the University of Western Australia, and Konrad Beyreuther, a protein-sequencing expert from the University of Cologne in Germany. Masters began telling Beyreuther about a human disease that featured plaques like those seen in scrapie and seemed to be very common. It was called Alzheimer's disease.

"Until then I had never heard of Alzheimer's disease," Beyreuther recalls.

It is easy to forget how recently Alzheimer's disease entered the public consciousness. For many decades after it first appeared in the medical literature, the term referred only to an obscure, early onset form of dementia. What we now know as common, late-onset Alzheimer's was then called 'senile dementia' — and it was so prevalent among the elderly that it hardly seemed worth classifying as a disease (see 'A problem for our age', page S2).

## THE MYSTERY PROTEIN

It is also easy to forget that at the dawn of Alzheimer's research, the disease was suspected of being a prion disease — we tend to think of the connection between prions and Alzheimer's as being much more recent. In late 2010, for example, a team led by Mathias Jucker at the University of Tübingen, Germany, reported that they could, in essence, transmit Alzheimer's-type brain pathology in a prion-like manner by injecting Alzheimer's brain matter into the bodies of mice (Eisele, Y. S. *et al. Science* 330, 980–982, 2010). Such findings have contributed to a major rethink of the cause of Alzheimer's disease. But this rethink is partly a renaissance because, as the story of Masters and Beyreuther's early interest in Alzheimer's reminds us, the prion connection is not new. "It was there at the beginning," says Masters.

As Masters knew in 1984, autopsies of Alzheimer's patients revealed brain plaques

resembling those seen in scrapie, often surrounded by dying neurons and their twisted axons and dendrites. When doused with Congo red, a standard pathology stain, and illuminated with polarized light, the Alzheimer's plaques — just like scrapie plaques — displayed an apple-green shimmer, a prismatic sign of the hydrogen bonds that held their fibrils tightly together. Protein aggregates that had this peculiar property were called amyloids.

Earlier that year George Glenner, an amyloid researcher at the University of California, San Diego, reported isolating a small protein from amyloid deposits in brain blood vessels in people with Alzheimer's disease. Was the protein embedded in Alzheimer's brain plaques the same as the one in Glenner's vascular deposits? Or was it more like the scrapie protein?

Masters and Beyreuther, at their dinner in Scotland, agreed to collaborate to find out, and their partnership probably did more than any other to launch modern Alzheimer's research and its central idea: the amyloid hypothesis.

Masters had already painstakingly purified a quantity of Alzheimer's amyloid, in a process akin to bomb-grade uranium enrichment. When a sample arrived in Germany, Beyreuther and his colleagues broke it down with formic acid and sifted through

**NATURE.COM**  
research the role  
of amyloid- $\beta$  in  
Alzheimer's disease  
[go.nature.com/qpcyez](http://go.nature.com/qpcyez)

the debris to find the smallest stable protein. This turned out to be a tiny peptide of roughly 40 amino acids, and Masters and Beyreuther called it A4. Sequencing A4 showed that it was not the scrapie protein, or indeed anything like it, but was essentially the protein Glenner had isolated from blood vessels.

Beyreuther's team quickly determined that A4 is a fragment of a much larger neuronal protein, amyloid precursor protein (APP). They found the gene that encodes APP on chromosome 21. This was a big clue, as people with Down's syndrome, who have an extra copy of chromosome 21, were known to develop Alzheimer's-like brain plaques by 40 years of age. The overproduction of APP and A4 was now revealed as the likely reason for the plaques in Down's syndrome — and probably in Alzheimer's disease too.

### TOO MUCH AGGREGATION

Other Alzheimer's investigators readily pursued the APP lead. But three other important clues from this initial burst of research by Beyreuther and Masters would be almost entirely overlooked for most of the next decade.

The first was an observation by Beyreuther about the forms of A4 in different solvent mixes. He noted the presence of stable clusters, or oligomers, made of two, four or more copies of A4. So strong was the peptide's tendency to form these oligomers that in certain solutions, dimers made of two copies of A4 were more prevalent than monomers.

The second clue was that full-length A4 is extremely prone to aggregate. After obtaining the full A4 sequence, Beyreuther began to synthesize various lengths of it in his lab, including a series that started at the 42nd (and terminal) amino acid of its longest variant and worked towards the opposite end. "When we came close to the end of the peptide and took it off the resin, we saw it getting aggregated," he remembers. "I thought 'Mein Gott, it's snowing!' It was aggregating so quickly. It was horrible."

The third clue came after Beyreuther and Masters raised the first antibodies to A4 and used them to detect amyloid deposits with unprecedented sensitivity in autopsied brains. The deposits were much more extensive than anyone had realized and were almost always present in people older than 80 years of age. In younger brains the plaques tended to be sparser and more diffuse, but they were still detectable in about 20% of cognitively normal people who had died in their fifties. The implication was that Alzheimer's disease is almost inevitable, with plaques beginning to form in the brain three decades before symptoms develop. "I thought that was amazing," says Beyreuther.

### THE AMYLOID HYPOTHESIS

By the end of the 1980s, Beyreuther and Masters had largely completed their discovery work on A4. Other scientists, mostly from the United States, took the lead on Alzheimer's research,



Like prions (above), amyloid- $\beta$  might spread in an infectious manner within tissues.

and one of their first acts was to rename the A4 protein amyloid- $\beta$ , where the  $\beta$  referred to the classic  $\beta$ -sheet molecular structure of amyloids. They also put much less emphasis on the original prion connection. "Some of these young guys who came after us didn't seem to know what a prion was," says Masters.

Even so, they seemed to move swiftly towards an understanding of how amyloid- $\beta$  causes Alzheimer's disease. In the early and mid-1990s, *in-vitro* studies indicated that amyloid- $\beta$  becomes toxic to neurons when it begins to aggregate. Genetic studies of families with early onset Alzheimer's disease detected mutations within the gene that encodes APP, and analysis of one of these mutant APP genes found that it causes a sevenfold overproduction of amyloid- $\beta$  (see 'Finding risk factors', page S20). Transgenic mice that overproduced human APP and amyloid- $\beta$  developed plaques resembling those seen in Alzheimer's disease, and their behaviour in standard tests suggested some cognitive deficits. The amyloid hypothesis seemed straightforward: when the amyloid- $\beta$  concentration in the brain becomes too high, the protein aggregates into fibrils and plaques, and begins killing neurons.

It eventually became clear that the situation was not quite that simple. Further genetic studies showed that familial, early onset Alzheimer's is usually caused not by the overproduction of total amyloid- $\beta$ , but by the relative overproduction of a less common variant of amyloid- $\beta$  known as A $\beta$ 42, the full-length, 42-amino-acid variant whose extreme proneness to aggregation had so alarmed Beyreuther.

The A $\beta$ 42 findings were still consistent with the plaque hypothesis, particularly once researchers recognized in the mid-1990s that the variant in most plaques is A $\beta$ 42. The problem was that mouse models with an over-dose of A $\beta$ 42 — like the first Alzheimer's mouse

models that overexpressed APP — lacked the heavy neuronal losses and cognitive decay associated with the human disease. "These models have some cognitive decline, but it's not as much as a person with full-blown Alzheimer's disease, by any stretch," says Harvard neurologist Bruce Yankner, a long-time Alzheimer's researcher.

Some researchers suspected that mice, with their small brains and short lives, cannot accurately model such a slow-burning, big-brain disease. But another possibility, which gained currency in the late 1990s, is that amyloid- $\beta$  plaques are not the real drivers of dementia. Autopsy studies showed, for example, that the progress of Alzheimer's dementia does not correlate well with the development of plaques. As Beyreuther and Masters had initially observed, the plaques become dense in the brain long before any signs of cognitive decline.

Unfortunately, the major pharmaceutical companies had already placed their bets on the amyloid- $\beta$  plaque hypothesis, and numerous drug-development programs would go on to fail in clinical trials. But in the meantime, a small group of researchers had begun to develop a new hypothesis that encompassed Alzheimer's and a variety of other amyloid-forming diseases.

### OLIGOMERS REVISITED

The genetic evidence made it almost certain that the aggregation of amyloid- $\beta$  somehow leads to Alzheimer's disease. The fibrils in plaques were the most obvious type of aggregate, and therefore the most obvious suspect. Only after the plaque hypothesis began to fail did researchers return to the other aggregates: the amyloid- $\beta$  oligomers first seen by Beyreuther and his colleagues in Cologne.

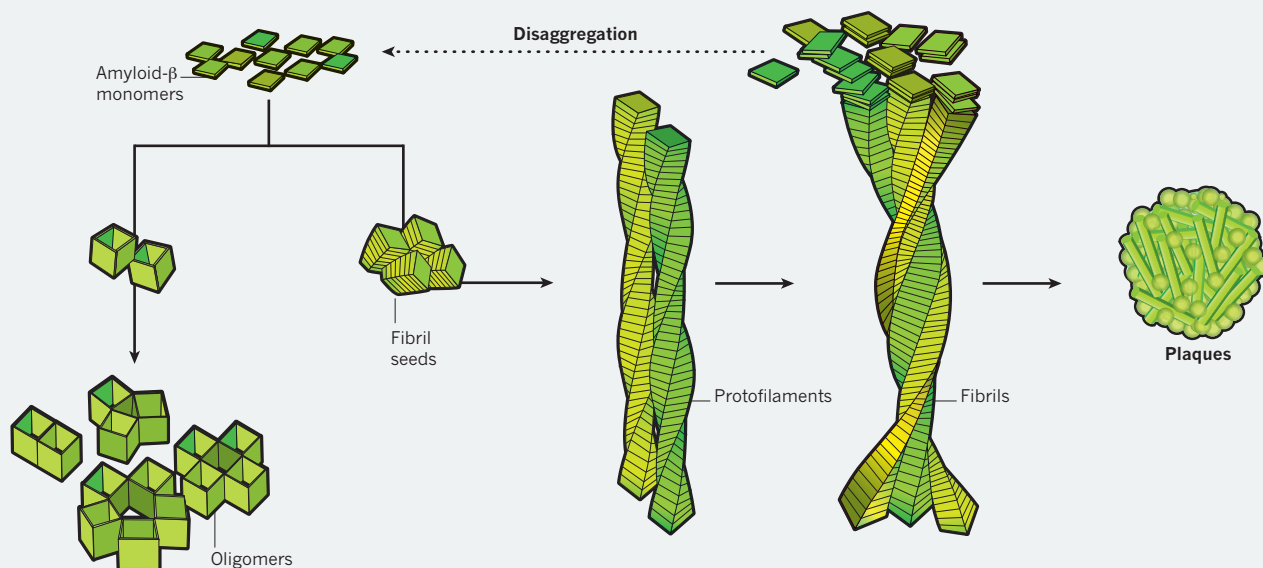
In the early and mid-1990s, Charles Glabe at the University of California, Irvine, and Dennis Selkoe at Harvard University reported finding oligomers in experiments with amyloid- $\beta$ . They saw them as briefly existing intermediates on the way to disease-causing fibrils, rather than fully fledged drivers of disease. But in 1998, William Klein's lab at Northwestern University in Evanston, Illinois, reported that oligomers could be the true culprits in Alzheimer's disease. When Klein's team added a chemical to a solution of amyloid- $\beta$  to stop it forming fibrils, the amyloid- $\beta$  instead formed oligomers, which then began to kill nearby neurons. At least some of this toxicity seemed to be the result of the oligomers weakening the synapses — the junctions between neurons — and impairing their ability to contribute to learning and memory (see 'Two pathways of aggregation'). Similar results soon followed from the Selkoe and Glabe labs, and in time mouse models also demonstrated oligomer toxicity.

In the 2000s, a new consensus began to emerge: that amyloid- $\beta$  fibrils are weakly toxic on their own, that they seem to provoke harmful inflammation, and that they are prone, especially



## TWO PATHWAYS OF AGGREGATION

Individual amyloid- $\beta$  peptides, which are produced normally by neurons, can assemble in at least two ways. One pathway leads to insoluble, plaque-forming fibrils. The other pathway leads to soluble oligomers, which are small enough to enter synapses. These oligomers are suspected to be the main toxic species in Alzheimer's disease.



when their plaques become especially dense, to slough off soluble amyloid- $\beta$  that can then reform into oligomers. But in this model, amyloid- $\beta$  oligomers are the more worrisome neurotoxins. Indeed, the amyloid- $\beta$  fibrils are probably protective to the extent that they trap aggregating amyloid- $\beta$  in a less harmful form.

Amyloid- $\beta$  oligomers are now thought to exert their harmful effects by binding directly to the membranes of neurons, or to specific receptors — the insulin and NMDA glutamate receptors are suspects — needed for neuronal signalling. But if amyloid- $\beta$  oligomers were merely toxic to neurons, they might never overwhelm the clearance mechanisms of the brain and cause disease. To do that, they seem to need another deleterious property that is associated with prions: infectiousness.

### PRIONS REVISITED

The idea that Alzheimer's might be a prion disease was first suggested in 1984 by the future Nobel laureate Stanley Prusiner of the University of California at San Francisco. His idea was widely dismissed after amyloid- $\beta$  was found to be different from scrapie protein. But by the mid 2000s, it was clear that Prusiner had been essentially correct. Both amyloid- $\beta$  and prion-disease proteins could fall into a state that was both toxic and self-replicating.

Prusiner, who was also at the dinner in Scotland with Masters and Beyreuther, was apparently wrong about the replication mechanism. He had initially proposed that an infectious prion is a protein monomer with

a misfolded shape that can induce the same misfolding in normal versions of the protein.

But as the chemist Peter Lansbury, then at Massachusetts Institute of Technology, showed in a series of *in-vitro* experiments in the mid-1990s, the key self-replicator in prion diseases and Alzheimer's disease appears to be an oligomer, not a monomer. Once formed, the oligomer becomes a template, or 'seed', that attracts new monomers, and aggregation around that nucleus proceeds rapidly. "This is one of those nonlinear phenomena in which small changes can have big effects," says Lansbury, now chief scientist at Link Medicine, a biotechnology company in Cambridge, Massachusetts.

One type of nucleus would serve as a template for new oligomers. Another would seed ever-lengthening fibrils. Lansbury showed that this initial nucleation event happens faster with a particularly sticky stretch of amino acids found on both prion proteins and A $\beta$ 42. Adding this stretch from A $\beta$ 42, or even adding full-length A $\beta$ 42, can trigger the runaway aggregation of all the amyloid- $\beta$  in the vicinity. Beyreuther's snow metaphor

*Alzheimer's disease is almost inevitable, with plaques beginning to form in the three decades before symptoms develop.*

was apt: a similar nucleation phenomenon lies at the heart of ice and snow crystallization.

More recently, Jucker and others have shown

that brain matter containing amyloid- $\beta$  from Alzheimer's patients can nucleate plaques in mice. Amyloid- $\beta$  is less hardy than prion proteins and so is much less likely to jump from one person to another, but it does seem to spread in an infection-like manner within tissues. "I was away from amyloid- $\beta$  research for years, but I've become interested again since Jucker showed that the stuff is infectious," says Beyreuther, who is now at the University of Heidelberg.

Similar infectious properties have been observed for aggregates of tau protein, which appear in Alzheimer's-affected cells late in the disease, as well as for  $\alpha$ -synuclein protein in Parkinson's disease. Researchers suspect that numerous other amyloid-linked diseases feature the same toxic, oligomeric mechanisms and involve a slow spread of pathology starting in the regions of the brain most vulnerable to the disorder. "We know, for example, that people who present with Parkinson's motor signs are almost always going to have Parkinson's dementia 20 years later," says Lansbury. In contrast, Alzheimer's disease affects memory and cognition quite early on.

In principle, according to Beyreuther, there could be protein structures in our food, air and water that get into the brain and promote disease-causing spirals of protein aggregation "like the little bit of dust that seeds the ice crystals in the windows," he says. "If that's true, then we are in trouble." ■

**Jim Schnabel** is a science writer based in Miami, Florida.



## PERSPECTIVE

# Prevention is better than cure

Attempts to reduce amyloid- $\beta$  in the brain have yet to show clinical benefits. Starting treatment early is the best hope, says **Sam Gandy**.

In 1907, Bavarian psychiatrist Alois Alzheimer published his description of a delusional woman who had slowly lost all cognitive function and died at 55 years of age. For decades thereafter, because of the patient's age, 'Alzheimer's presenile dementia' was considered a rare disease of mid-life. In the 1970s, neuropathologists realized that 'senile dementia' was indistinguishable from the disease described by Alzheimer<sup>1</sup>. The clinical picture of progressive brain dysfunction in association with the post-mortem brain pathology of extracellular amyloid deposits ('plaques') and intraneuronal 'tangles' was renamed 'Alzheimer's disease', and its definition was broadened to include the major cause of dementia that we know today.

The accumulation of amyloid plaques is the key distinguishing feature of Alzheimer's disease, and research in the late 1980s identified the major plaque protein as the amyloid- $\beta$  peptide. In the 1990s, scientists connected amyloid- $\beta$  pathology with genes encoding three proteins: the amyloid- $\beta$  precursor protein (APP), which is cleaved by  $\gamma$ -secretase to create amyloid- $\beta$ , and two forms of presenilin (presenilin 1 and presenilin 2), which are involved in amyloid- $\beta$  generation<sup>2</sup> (see 'Finding risk factors, page S20). Together, mutations in these genes account for about 3% of Alzheimer's disease, and insertion of one of these mutations into the mouse genome created the first transgenic mice to form amyloid plaques.

But what about the other 97% of Alzheimer's disease? Parsimony would dictate that amyloid- $\beta$  pathology might be central to all forms of Alzheimer's disease<sup>3</sup>. Less than five years after the plaque-forming mice were developed, the first amyloid- $\beta$ -lowering drugs and vaccines were identified. However, initial human trials of the drugs were uninterpretable because they included no methods for measuring how much amyloid- $\beta$  is in the brain. That challenge was overcome in 2004 using positron emission tomography to identify amyloid plaques<sup>4</sup>.

And that's where the field stood for six years. Then, in 2010, researchers at the University of Turku in Finland showed that



The process of Alzheimer's can begin 20 years before the mind shows signs of cognitive loss.

immunotherapy with an anti-amyloid- $\beta$  monoclonal antibody lowered plaque burden by around 25%. Disappointingly, however, the treatment — an infusion of antibody every three weeks for 1.5 years — brought no cognitive benefit for the patient<sup>5</sup>.

Why did this immunotherapy reduce brain plaques but fail to halt cognitive decline? Perhaps 1.5 years is not long enough or perhaps a 25% reduction in plaque burden is insufficient. To allow for this possibility, immunotherapy trials are continuing and results are expected in 2013. Another possibility is that the monoclonal antibody used might not recognize the most important neurotoxic conformation(s) of amyloid- $\beta$ . Dozens of different monoclonal antibodies, as well as intravenous immunoglobulin, are now in clinical trials, in the hope that one or several will recognize and neutralize the most neurotoxic forms of amyloid- $\beta$ .

There is at least one more possible interpretation of the Turku study: that therapy to lower amyloid- $\beta$  levels will never succeed in symptomatic patients. Brain imaging data from presymptomatic individuals who carry presenilin 1 mutations show that plaque accumulation starts 10–20 years before

clinical symptoms appear<sup>6</sup>. So if subjects enter trials at the first sign of cognitive impairment, they might already harbour substantial quantities of neurotoxic amyloid- $\beta$ .

The best hope for therapies aimed at amyloid- $\beta$  levels, therefore, is to dose prophylactically to stop it building up in the first place. A diagnostic category of 'presymptomatic Alzheimer's disease' was recently proposed for subjects with strong biomarker-based evidence of disease but who are cognitively intact<sup>7</sup>. Nevertheless, in the absence of a perfect test for predicting who will develop Alzheimer's disease and when, prevention trials are highly daunting with regard to cohort size, trial duration and cost. The most obvious place to start is with carriers of presenilin 1, presenilin 2 or APP mutations, where disease risk and timing of onset are highly predictable. The Dominantly Inherited Alzheimer Network has been founded to identify carriers of pathogenic mutations worldwide and enrol them into prevention trials with amyloid- $\beta$ -lowering agents.

The amyloid- $\beta$  odyssey of the past 25 years has shown that conquering Alzheimer's disease is not a matter of removing amyloid- $\beta$  plaques from the brain *post hoc*. But the role of amyloid- $\beta$  must be resolved, and our quest for effective interventions must be seen through to a successful end. Alzheimer's disease is already a problem for the healthcare systems of Western countries and is a growing threat to developing nations.

The best argument for sticking with strategies to lower amyloid- $\beta$  levels is that safe, effective compounds are within reach. Perfecting the selection of subjects and the timing of intervention could delay the onset of Alzheimer's disease substantially. A century of effort has brought us to a rational model for how Alzheimer's disease might begin, and we should not be discouraged by the prospect of another decade or two of work to settle the amyloid- $\beta$  issue and ultimately, we hope, defeat the illness. Prophylactic intervention to lower amyloid- $\beta$  levels is now the best hope. ■

**Sam Gandy** is neurology and psychiatry professor and associate director of the Alzheimer's Disease Research Center at Mount Sinai School of Medicine and the James J. Peters VA Medical Center in New York. e-mail: samuel.gandy@mssm.edu

1. Terry, R. D. & Katzman, R. *Ann. Neurol.* **14**, 497–506 (1983).
2. Bertram, L., Lill, C. M. & Tanzi, R. E. *Neuron* **68**, 270–281 (2010).
3. Gandy, S. J. *Clin. Invest.* **115**, 1121–1129 (2005).
4. Klunk, W. E. et al. *Ann. Neurol.* **55**, 306–319 (2004).
5. Rinne, J. O. et al. *Lancet Neurol.* **9**, 363–372 (2010).
6. Knight, W. D. et al. *Brain* **134**, 293–300 (2011).
7. Sperling, R. A. et al. *Alzheimers Dement.* **7**, 280–292 (2011).

The author declares competing financial interests: go.nature.com/hdiuds



ILLUSTRATION BY GRACIA LAM

## PREVENTION

# Activity is the best medicine

*Can exercise, social interaction and the Mediterranean diet really help to keep the cognitive decline of Alzheimer's disease at bay?*

BY SARAH DEWEERDT

Rhumba. Lindy hop. Cha-cha. Ballroom dancing may not be the first preventive treatment for Alzheimer's disease that springs to mind, but it is an ideal prescription for those concerned about their declining memory. In fact, says Perminder Sachdev, a neuropsychiatrist at the University of New South Wales in Sydney, Australia, dancing has a perfect blend of elements that help stave off dementia. "There's cognitive activity, there's also physical activity, and there's social interaction as well."

A healthy Mediterranean-style diet is also thought to be protective — so that dance class could be topped off with a big Greek salad and a glass of red wine.

Over the past decade, epidemiological studies have shown that exercise, intellectual activity, social relationships and a healthy diet all lead to a lower risk of dementia. Such findings have to be interpreted with caution, however, because many researchers are sceptical about

the benefits, and because withdrawing from social relationships and other activities can be an early symptom of dementia, not just a risk factor for it.

Even so, "we have enough suggestive observational data now from several studies" to conclude that lifestyle factors are important in Alzheimer's disease, much as they are in cardiovascular disease, says Ronald Petersen, director of the Alzheimer's Disease Research Center at the Mayo Clinic in Rochester, Minnesota.

## FIT AND HEALTHY

The task now is to move from lifestyle factors to interventions — to find out how much exercise, what kind of intellectual activity and at what stage each could influence the course of the disease. "We need to do more [clinical trials] where we actually intervene" with cognitive activity, training programmes and exercise, and with an appropriate control group, Petersen says.

Some of these trials are already under way. For example, in the Fitness for the Aging Brain Study, researchers<sup>1</sup> in Australia recruited 170 people who were worried that their memory had deteriorated or who had mild cognitive impairment (MCI), a condition that is considered a precursor to Alzheimer's disease. The researchers assigned half of the participants to a six-month exercise programme, either walking or doing other aerobic exercise for 50 minutes, three times a week. The other half, in the control group, carried on with their usual level of activity.

After six months, those in the exercise group slightly improved their scores on the cognitive section of the Alzheimer's Disease Assessment Scale (ADAS-Cog), a series of short memory, language and reasoning tests, whereas control subjects declined at a rate consistent with normal ageing. What's more, the exercise had lasting effects, leading to better scores 12 months after the programme ended.

ADAS-Cog is commonly used in clinical trials of Alzheimer's disease drugs, so the researchers were able to compare the effects of exercise with those of drugs called acetylcholinesterase inhibitors, which reduce the breakdown of the neurotransmitter acetylcholine. For people with MCI, regular exercise "can help your brain more than taking the medication that is currently available for Alzheimer's disease", says one of the researchers, Nicola Lautenschlager, who studies geriatric psychiatry at the University of Melbourne.

## PHYSICAL CHANGES

How does this connection between body and mind work? Studies in rodents have suggested at least two different mechanisms<sup>2</sup>. First, exercise increases the activity of an enzyme called neprilysin that metabolizes amyloid- $\beta$  — the protein that makes up the characteristic plaques of Alzheimer's disease — and might

➔ **NATURE.COM**  
more research on the  
links between life-  
style and dementia  
[go.nature.com/ivcdp8](http://go.nature.com/ivcdp8)



help clear it from the brain. Physical activity also turns on the production of brain chemicals such as nerve growth factors, which promote the formation of nerve cells and the connections between them. This process is thought to make the brain better able to cope despite the pathological changes of Alzheimer's disease.

In the past few years, the development of biomarkers (see 'Warning signs', page S5) that can indicate Alzheimer's-related brain changes in living people have allowed researchers to explore more fully the mechanisms of the mind-body connection. For example, one study this year of 120 sedentary but healthy older adults without any memory problems assigned half the participants to a 3-days-a-week programme of physical exercise<sup>3</sup>. After a year, researchers performed magnetic resonance imaging (MRI) on several brain areas, including the hippocampus, the brain structure responsible for memory formation.

In older adults, the hippocampus typically shrinks by 1–2% each year, and this is what happened in the control group. But in the exercise group, the volume of the hippocampus actually increased by 2%. "That's probably millions of cells," says research team member Kirk Erickson, a psychologist at the University of Pittsburgh, Pennsylvania. With one year of exercise, "we are in essence rolling back the clock by one to two years".

## BRAIN TRAINING

Another report<sup>4</sup>, also published this year, suggests that similar mechanisms are at work when people exercise their brains. Canadian researchers used functional MRI to analyse brain activity in 15 people with MCI. After a one-week programme designed to teach the participants new memory strategies, there was activation in several additional brain regions during memory tests, suggesting that intact areas of the brain were able to take over from damaged areas. The participants also scored better on the tests.

Many studies of cognitive stimulation and dementia make use of computer games designed to boost mental skills. Although such 'brain training' interventions do not generally make healthy people smarter, they produce positive results in people with Alzheimer's disease and related conditions. One 2006 trial funded by the US National Institutes of Health showed that brain training can counteract some of the cognitive decline expected with ageing<sup>5</sup>. In that study — known as Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE) — people over 65 years of age who did a five- to six-week brain training programme focusing on memory, reasoning or speed of processing skills were better at these skills than control participants even five years later.

Computerized brain training programs are popular among researchers because these

interventions are controllable and predictable, especially compared with intellectual pursuits in the real world. But this doesn't mean that people need to play computer games to stay mentally agile, says Sachdev. Instead, he argues, people are likely to benefit from any intellectual pursuit that both requires effort ("something where you challenge your brain") and is enjoyable ("so that you can sustain it"). That could mean anything from taking up the clarinet to doing Sudoku puzzles.

## FOOD FOR THOUGHT

Meanwhile, other lifestyle factors that can modify the risk of Alzheimer's disease are continuing to emerge through epidemiological research. These types of studies, involving observation of thousands of people and their habits, underpin our knowledge about the Mediterranean diet, which includes a relatively high consumption of fruits, vegetables, whole grains and olive oil, relatively low consumption of red meat and saturated fat, and a glass of red wine with dinner.

Eating these foods has already been shown to reduce the risk of cardiovascular disease, hypertension and diabetes. In the past few years, three independent epidemiological studies conducted in New York<sup>6</sup>, Chicago<sup>7</sup> and Bordeaux, France<sup>8</sup>, have shown that those who eat mostly Greek peasant food also stay the sharpest mentally. "There has been converging evidence that adherence to such a diet is related to lower risk of cognitive decline or Alzheimer's disease," says Nikolaos Scarmeas, a neurologist at Columbia University in New York.

A team of Columbia University researchers including Scarmeas asked 1,880 New Yorkers detailed questions about their eating habits, then studied them for an average of five-and-a-half years. They found that the people with the most Mediterranean diet have up to a 40% lower risk of developing Alzheimer's disease than those who eat less Mediterranean food<sup>6</sup>. Results like these are so promising that several groups around the world are planning randomized trials of the Mediterranean diet as a way of preventing Alzheimer's disease.

Evidence that social engagement helps to prevent dementia also comes primarily from observational studies. For example, among more than 6,000 people aged 65 or older in Chicago, those with the most extensive social networks and the highest levels of social engagement have the lowest rates of cognitive decline<sup>9</sup>.

It can be difficult to measure people's level of social engagement and it is even harder to design randomized trials to investigate it. Disentangling the effects of social engagement from those of other lifestyle elements is far from straightforward. Still, social engagement is a form of intellectual engagement, argues Linda Teri, professor of psychosocial and community health at the University of Washington in Seattle. Teri has designed

programmes to encourage physical activity and social connections in people with MCI and dementia. "When we are with other people, we are listening to the conversation, we're tracking ideas, we're forming our own ideas," she says. "We're actually engaging in quite a bit of cognitive skills."

So people who exercise in groups may benefit from both the social stimulation and the physical activity. For example, consider Erickson and colleagues' research into exercise and brain changes in healthy older people<sup>3</sup>. Instead of aerobic exercise, the control group met three times a week to do stretches. This did not increase the size of their hippocampus, but it did improve their scores on a simple computerized test of memory, similar to the improvements in the exercise group. Erickson suggests that this social stimulation benefits other parts of the brain that the study did not measure.

## A LITTLE BIT BETTER

In some parts of the research community, the argument that lifestyle can help to delay Alzheimer's disease is a tough sell. Last year, the US National Institutes of Health organized a consensus panel on preventing Alzheimer's disease. It concluded that it is too soon to tell whether lifestyle changes — or any other prevention strategy — can affect the development or the course of Alzheimer's disease.

Even those who are more bullish about the evidence say that lifestyle changes are likely to have only a limited benefit. But because Alzheimer's disease develops late in life, even small changes in risk or slight delays in the development of symptoms could greatly reduce the burden of disease, as people would be more likely to die from other causes before becoming mentally impaired.

As Erickson says: "If we can at least prevent some of the normal age-related decline from happening, even if it doesn't eliminate the risk — if it just reduces the risk of developing Alzheimer's disease or makes the quality of life a little bit better — I think we've gone a long way." ■

*Sarah DeWeerd is a science writer based in Seattle, Washington.*

1. Lautenschlager, N. T. *et al.* *J. Am. Med. Assoc.* **300**, 1027–1037 (2008).
2. Savica, R. & Petersen, R. C. *Psychiatr. Clin. N. Am.* **34**, 127–145 (2011).
3. Erickson, K. I. *et al.* *Proc. Natl Acad. Sci. USA* **108**, 3017–3022 (2011).
4. Belleville, S. *et al.* *Brain* **134**, 1623–1634 (2011).
5. Willis, S. L. *et al.* *J. Am. Med. Assoc.* **296**, 2805–2814 (2006).
6. Scarmeas, N. *et al.* *J. Am. Med. Assoc.* **302**, 627–637 (2009).
7. Tangney, C. C. *et al.* *Am. J. Clin. Nutr.* **93**, 601–607 (2011).
8. Féart, C. *et al.* *J. Am. Med. Assoc.* **302**, 638–648 (2009).
9. Barnes, L. L., Mendes de Leon, C. F., Wilson, R. S., Bienias, J. L. & Evans, D. A. *Neurology* **63**, 2322–2326 (2004).



## VACCINES

# Chasing the dream

*After a decade of disappointments, hopes for a successful Alzheimer's vaccine that ameliorates symptoms and ultimately prevents the disease are rising again.*

BY JIM SCHNABEL

Using the formidable powers of the immune system to attack one of the body's own proteins seems like a risky approach. But this is what nearly all vaccines, or immunotherapies, against Alzheimer's disease aim to do. Their target is amyloid- $\beta$ , a tiny protein produced by neurons. Scientists do not know what function amyloid- $\beta$  evolved to have in its ordinary, free-floating form. But they do know that it is unusually prone to sticking to copies of itself, and that this aggregation process seems to be the principal trigger for Alzheimer's disease.

The first vaccine against Alzheimer's disease — Dublin-based Elan Pharmaceuticals' AN-1792 — was based on a particularly aggregation-prone form of amyloid- $\beta$  known as A $\beta$ 42. In mice that had Alzheimer's-like deposits, or 'plaques', of amyloid- $\beta$  in their brains, it seemed enormously promising: it provoked a storm of anti-amyloid- $\beta$  antibodies that dissolved the plaques in older mice and stopped plaques from forming in younger ones. But in humans, AN-1792 was a disaster. Elan halted its first large clinical trial in 2002, after patients developed meningoencephalitis, an inflammation of the brain and its membranes that was apparently caused by rogue immune cells<sup>1</sup>.

Most subsequent efforts have fared little better. Milder, second-generation active vaccines against amyloid- $\beta$  are still in clinical trials, but many researchers suspect that these will not be strong enough to provoke a sufficient antibody response in elderly patients with weak immune systems. Passive vaccine infusions of lab-grown anti-amyloid- $\beta$  antibodies are meant to get round this problem, but they haven't performed well in clinical trials.

"We in the field have had to look back and say, what did we do wrong?" says Norman Relkin, a neurologist at Weill Cornell Medical College, part of Cornell University in New York.

But despite these disappointments, there are hints of clinical success from a surprising direction — one that could lead to a better understanding of Alzheimer's disease and to therapies and preventives that really work.

The vaccine that has raised some researchers' hopes is a mix of antibodies pooled from donated human blood. Known as intravenous immunoglobulin (IVIg), it has long been marketed as a general booster for antibody-based immunity in people who lack it for genetic reasons, and as a moderator of some rare autoimmune conditions.

The idea of using IVIg to treat Alzheimer's disease occurred to Relkin and his colleague Marc Weksler after they found, in 2002, that people with Alzheimer's disease have lower levels of anti-amyloid- $\beta$  antibodies in their blood than cognitively normal people of the same age. They decided to set up a small, 6-month study of IVIg in eight of Relkin's patients. "The concept simply was to give back these antibodies, since IVIg is derived from the plasma of young individuals who tend to have higher levels," Relkin says.

The results were surprisingly good: six patients improved their cognitive scores, and a seventh stabilized. In a larger trial of 24 patients, Relkin again found signs that IVIg was working: the eight-person placebo group worsened as expected, but nearly all the 16 treated patients improved moderately on both cognitive and quality-of-life measures over the first 6 months (ref. 2). Their improvements were roughly equivalent to turning back the clock by 6–18 months. What's more, they stayed at those levels for as long as the treatment continued — more than two years in some cases.

## INJECTION OF REALISM

The results of small trials often fail to hold up in larger trials. But Relkin's results have inspired some optimism — and some off-label prescribing of IVIg for Alzheimer's disease — because the improved cognitive and behavioural scores were dose dependent and have been backed up by changes in biological markers, including lower levels of amyloid- $\beta$  in cerebrospinal fluid and reduced brain shrinkage. In fact, Relkin says, brain shrinkage is "towards the normal range in individuals who got the best dose, which is a very provocative finding".

The US National Institute on Aging, along with Baxter BioScience of Deerfield, Illinois, one of several producers of IVIg, is sponsoring a follow-up trial in 400 individuals with Alzheimer's disease. The results could be ready by the end of 2012. If the trial is successful,

it could lead to the first Alzheimer's therapy approved by the US Food and Drug Administration that modifies the disease, rather than just treats the symptoms.

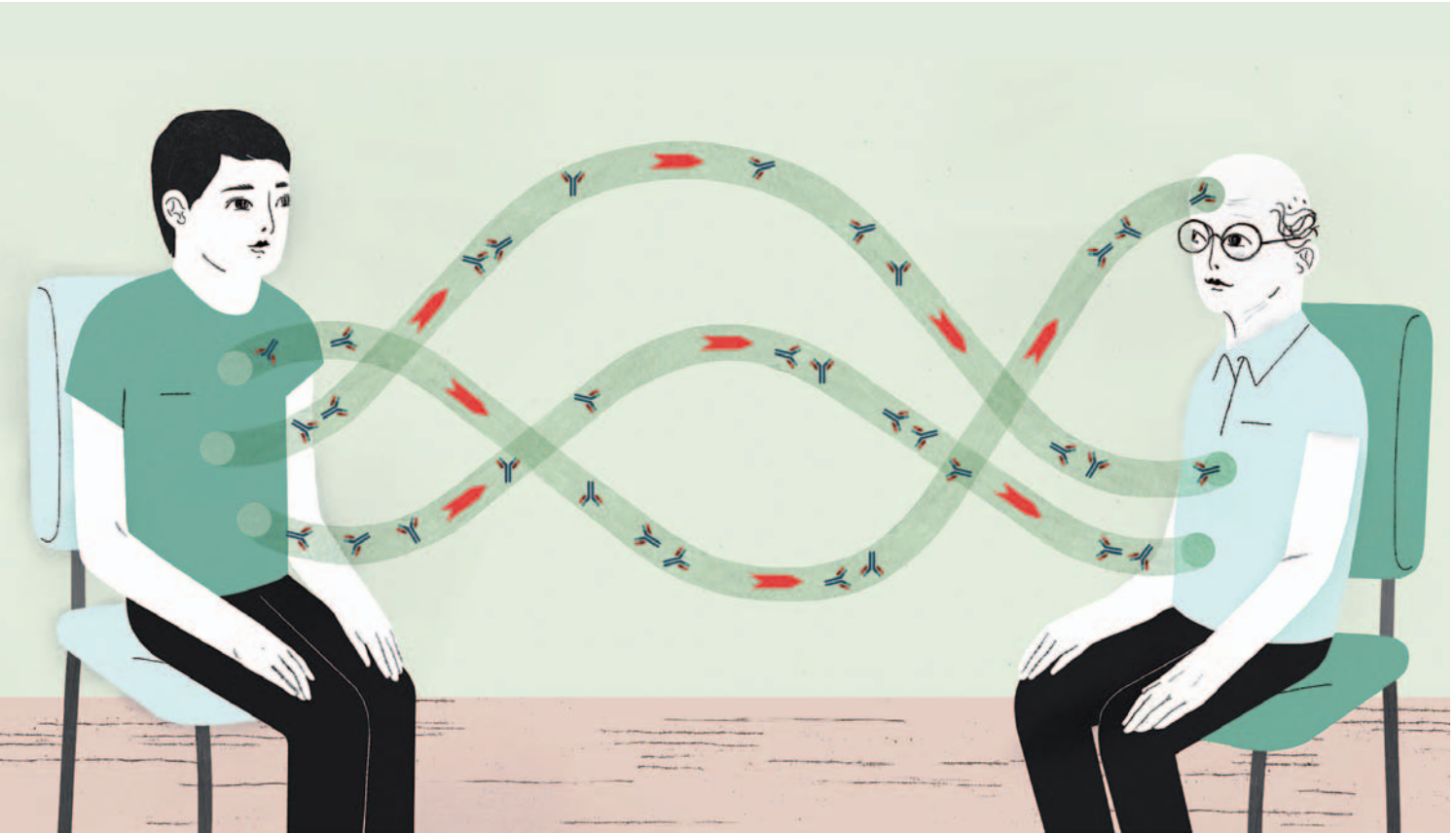
But this would not be the end of the Alzheimer's story, merely the end of the beginning. IVIg has several shortcomings. First, those who seemed to benefit from treatment had only modest gains. "I have not seen anyone re-enrol in adult education classes," says Weksler. Second, there seems to be a limited window of time when the therapy is effective. In the small trials carried out so far, the patients who started IVIg treatment later in the disease course seemed more likely to keep worsening.

There are also problems of cost and availability. IVIg is infused at high doses every two weeks in these studies, and patients might need them for the rest of their lives, at a cost of thousands of US dollars per infusion. Worse still, the production capacity for blood products from human donors is limited, and demand for IVIg from Alzheimer's patients and their families would swiftly outstrip supply. "We need next-generation products that are easier to produce and are based on IVIg's mechanisms of action," says Relkin.

Unlike most other Alzheimer's vaccines, IVIg has several plausible mechanisms. Although some of its antibodies may keep aggregates of amyloid- $\beta$  in check, others may counter brain inflammation and reduce aggregates of tau protein, which also contribute to dementia. "You're talking about a complex disease that has many different pathological processes occurring either sequentially or in parallel," says Relkin. "So IVIg in this respect is ideally suited."

By contrast, AN-1792 and other big pharma Alzheimer's vaccines have aimed squarely at amyloid- $\beta$  in its natural, single-copy form, as well as in fibrils — the long, insoluble, plaque-making aggregates that show up prominently in the brain and cerebral blood vessels of Alzheimer's patients. The lack of success with these vaccines suggests that single-copy and fibril amyloid- $\beta$  might not be the best targets in patients who already have dementia.

So far, for all these vaccines, there has been only one published efficacy study: a phase II trial of bapineuzumab, Elan's passive anti-amyloid- $\beta$  antibody infusion. The beneficial effects of bapineuzumab seemed weak to non-existent and, even worse, at high doses it caused brain swelling and associated



microbleeds in some patients with heavy vascular amyloid- $\beta$  deposits<sup>3</sup>. Autopsy and brain imaging studies of selected bapineuzumab and AN-1792 recipients suggest that these vaccines can fail to slow the progress of dementia even when they succeed in reducing plaques of amyloid- $\beta$  in the brain<sup>4</sup>.

One reason for these disappointing results may be that the vaccines address only amyloid- $\beta$  and do nothing to counteract brain inflammation or tau aggregates. Another possibility is that they are less effective at clearing the small, soluble clusters of amyloid- $\beta$  known as oligomers, which are now seen as far more toxic than fibrils and which seem to promote the appearance of tau aggregates<sup>5</sup> (see 'Little proteins, big clues', page S12).

The short-term effects of IVIg could be due to its ability to clear amyloid- $\beta$  oligomers, Relkin says. "Studies have suggested that you can reverse signs of memory impairment in mouse models within 24 hours of giving anti-oligomer antibodies," he says. "It's wonderful that we have a potential therapeutic as well as something that is directing us towards new avenues, new mechanisms, in studying the problem."

### DREAM VACCINES

In the future, vaccines may also be used to treat people who have less advanced disease and so might get more benefit. "We're all moving towards the idea of treating patients with very mild dementia or even before they

develop symptoms," says Dennis Selkoe, a neurologist at Harvard Medical School and long-time Alzheimer's researcher.

"The ultimate dream is to be able to give people a vaccine when they're still in their 20s or 30s, to prevent the disease process from even starting," says Cynthia Lemere, a Harvard neurobiologist who tests active anti-amyloid- $\beta$  vaccines in monkeys.

Lemere, Selkoe and others believe that until dementia sets in, amyloid- $\beta$  is the main driver of disease. Even the existing vaccine candidates might work well in this presymptomatic phase by keeping amyloid- $\beta$ , in all its forms, within manageable levels.

Other researchers favour a universal Alzheimer's vaccine that leaves ordinary, single-copy amyloid- $\beta$  alone and instead targets structures found only on amyloid- $\beta$  aggregates, particularly oligomers and incipient fibrils. According to Relkin, the natural anti-amyloid- $\beta$  antibodies found in IVIg seem to target these shapes, rather than single-copy amyloid- $\beta$ .

"I see these as pathology-specific structures, so they're ideal targets," says Charles Glabe, an Alzheimer's vaccine researcher at the University of California, Irvine. "I think you'd have your best therapeutic effect this way, and the fewest side effects."

To elicit antibodies against these targets, Glabe and others have vaccinated animals with synthetic peptides that have the desired shapes but contain non-human amino-acid sequences, lowering the risk of autoimmune

reactions. These vaccines reduce brain pathology and improve memory-related behaviours in mouse models of Alzheimer's disease, just as broader anti-amyloid- $\beta$  vaccines do<sup>6</sup>. In principle, some of the aggregate-specific antibodies evoked by these vaccines would bind to aggregates of other disease-linked proteins, such as  $\alpha$ -synuclein in Parkinson's disease or prion proteins in Creutzfeldt-Jakob disease (CJD), so the same approach could be used against all such diseases.

So far, none of these third-generation vaccines has had the corporate backing to reach clinical trials, but that could change quickly. "If one of the existing vaccines shows a strong effectiveness profile in clinical trials, then I think interest will go way up," says Glabe. He would particularly welcome a success for IVIg, because it is widely believed to work on the same principle as an oligomer vaccine. "But investors tend to lump all immunotherapies together," he says, "so they rise and fall together even though they may have very different targets." ■

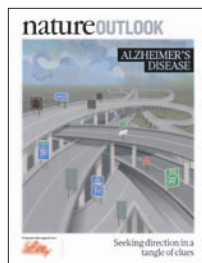
**Jim Schnabel** is a science writer based in Miami, Florida.

1. Gilman, S. *et al. Neurology* **64**, 1553–1562 (2005).
2. Hughes, R. A. C. *et al. Clin. Exp. Immunol.* **158** (Suppl 1), 34–42 (2009).
3. Salloway, S. *et al. Neurology* **73**, 2061–2070 (2009).
4. Rinne, J. O. *et al. Lancet Neurol.* **9**, 363–372 (2010).
5. Jin, M. *et al. Proc. Natl Acad. Sci. USA* **108**, 5819–5824 (2011).
6. Goñi, F. *et al. PLoS One* **5**, e13391 (2010).

# natureOUTLOOK

## ALZHEIMER'S DISEASE

14 July 2011 / Vol 475 / Issue No. 7355



COVER ART: NIK SPENCER

### Editorial

Herb Brody,  
Michelle Grayson,  
Apoorva Mandavilli,  
Tony Scully,  
Nick Haines

### Art & Design

Alisdair MacDonald,  
Barbara Izdebska

### Production

Karl Smart, Emilia  
Orviss, Leonora  
Dawson-Bowling,  
Stephen Russell

### Sponsorship

Gerard Preston, David  
Bagshaw, Yvette Smith

### Marketing

Elena Woodstock,  
Hannah Phipps

### Project Managers

Helen Anthony,  
Claudia Deasy

### Art Director

Kelly Buckheit Krause

### Magazine Editor

Tim Appenzeller

### Editor-in-Chief

Philip Campbell

### Editorial advisor

Marie-Thérèse Heemels

The fog that envelops so many people as they age, severing them from their memories and thus from their identity, used to be considered a normal part of growing old — along with sore joints, needing reading glasses and losing touch with popular music. However, to anyone who has seen a loved one slip behind the heavy curtains of what we now call Alzheimer's disease, the decline seems anything but natural. What kind of massive malfunction in the brain can send an alert, robust, witty person into a state of persistent confusion?

The theory that plaques of amyloid- $\beta$  in the brain trigger the disease has been called into question (page S12); Alzheimer's disease is a subtler foe. And without a handle on the disease's cause or genetic underpinnings (page S20), the developers of drugs (page S9) and vaccines (page S18) are working in a fog of their own. Moreover, Alzheimer's disease cannot be definitively diagnosed without an autopsy of the brain — at which point the information is rather academic, at least for that individual. So the search is intensifying for biomarkers — clues that indicate reliably whether a person who is still alive and healthy is destined for Alzheimer's disease (page S5).

The stakes are high. Alzheimer's disease is a drain not only on individuals and families, but also on societies, with the costs of care and lost productivity exceeding US\$300 billion per year, which will only increase with rising incidence. More people than ever are making it to old age, but dementia is the reward for 6 out of every 100 individuals who get past 60 years (page S2).

We can take some encouragement from the findings that there may be non-medical steps that people can take to ward off the disease (page S16) — and that the prescribed activities, such as dancing and playing games, are pleasant enough in their own right.

We thank Eli Lilly for the financial support that has made this *Outlook* possible. As always, *Nature* retains sole responsibility for all editorial content.

**Herb Brody**

*Supplements Editor, Nature Outlook.*

## CONTENTS

### S2 DEMENTIA

#### **A problem for our age**

An ageing global population means dementia is a rapidly growing disease

### S5 BIOMARKERS

#### **Warning signs**

New ways to track cognitive decline, from brain imaging to simple patient tests

### S8 PERSPECTIVE

#### **In search of biomarkers**

Neil S. Buckholtz

### S9 DRUGS

#### **A tangled web of targets**

Obscure pathology is keeping drug development in the dark

### S12 AMYLOID

#### **Little proteins, big clues**

Fibril clumps and tau tangles are the usual suspects, but are they guilty?

### S15 PERSPECTIVE

#### **Prevention is better than cure**

Sam Gandy

### S16 PREVENTION

#### **Activity is the best medicine**

A healthy lifestyle might ward off the symptoms of Alzheimer's

### S18 VACCINES

#### **Chasing the dream**

Setbacks and successes in the pursuit of disease-modifying treatments

### S20 GENETICS

#### **Finding risk factors**

A diverse range of genes are implicated in the disease, complicating the picture

## COLLECTION

### S23 Amyloid- $\beta$ and tau — a toxic pas de deux in Alzheimer's disease

L. M. Ittner & J. Götz

### S29 The ethics of informed consent in Alzheimer disease research

Scott Y. H. Kim

### S34 Recollection of lost memories

R. C. Malenka & R. Malinow

### S35 Gamma-secretase activating protein is a therapeutic target for Alzheimer's disease

He et al.

### S39 Clinical trials of disease-modifying therapies for neurodegenerative diseases: the challenges and the future

Anthony E. Lang

*Nature Outlooks* are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the *Nature Outlook Editorial guidelines* available at [http://www.nature.com/advertising/resources/pdf/outlook\\_guidelines.pdf](http://www.nature.com/advertising/resources/pdf/outlook_guidelines.pdf)

### CITING THE OUTLOOK

Cite as a supplement to *Nature*, for example, *Nature* Vol XXX, No. XXXX Suppl, Sxx–Sxx (2011). To cite previously published articles from the collection, please use the original citation, which can be found at the start of each article.

### VISIT THE OUTLOOK ONLINE

The *Nature Outlook Alzheimer's disease* supplement can be found at <http://www.nature.com/nature/outlook/Alzheimers/>

All featured articles will be freely available for 6 months.

### SUBSCRIPTIONS AND CUSTOMER SERVICES

For UK/Europe (excluding Japan): Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas — including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea: Nature Publishing Group — Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

### CUSTOMER SERVICES

Feedback@nature.com  
Copyright © 2011 Nature Publishing Group





More than 90% of Alzheimer's cases manifest in people over 65 years of age.

#### GENETICS

# Finding risk factors

*Uncovering genes that are linked with Alzheimer's disease can help researchers understand what causes the disease. But it's not easy.*

BY MICHAEL EISENSTEIN

Headlines trumpet the discovery of genes associated with Alzheimer's disease so often that one might think the genetic foundations of the disease must surely be mapped out in their entirety. Certainly for those who develop the early onset, or familial, form of the disease in late middle age, the lion's share of the blame can be attributed to three genes: *APP*, *PSEN1* and *PSEN2*. Each of these genes plays a role in producing amyloid- $\beta$ , the accumulation of which is widely thought to trigger the disorder's characteristic neurodegeneration.

However, more than 90% of Alzheimer's

cases are of the late-onset form, which typically manifests in people older than 65 years and seems to have a separate pool of genetic risk factors. Efforts to identify factors directly involved in the processing and accumulation of amyloid- $\beta$  have yielded at least a dozen candidate genes implicated in this form of the disease, but their roles are still unclear and their total contribution cannot account for the estimated 60–80% hereditary risk of late-onset disease<sup>1</sup>.

One factor — a common variant of the gene encoding apolipoprotein E (ApoE) — has come to dominate the Alzheimer's landscape<sup>2</sup>. Just one copy of this variant, called *APOE4*, increases disease risk fourfold; two copies

raise the risk about tenfold. "If you're going to try to predict who's going to get Alzheimer's, *APOE* is probably equivalent to the rest of the genes combined," says Gerard Schellenberg, director of the US-based Alzheimer's Disease Genetics Consortium.

Although *APOE* plays a leading role in the Alzheimer's story, it relies on a large supporting cast. Discovery of these other genetic players gained momentum with the rise of genome-wide association studies (GWAS). In this approach, researchers analyse millions of single nucleotide polymorphisms (SNPs) — variations scattered throughout the genome — in tens of thousands of affected and healthy individuals. By finding genomic changes that correlate with disease, they can uncover candidate genes or harmful mutations.

#### STATISTICAL POWER

Well over a dozen GWAS studies on Alzheimer's disease have been published, most of them from large consortia in Europe and the United States. Studies of this sort are often criticized for finding false positive associations, which cannot be replicated by other studies, and the early Alzheimer's studies were no exception. But later efforts analysed many more SNPs in the genomes of large populations of people with little overall genetic variability between them, increasing the statistical power and allowing scientists to identify variants in more than ten genes associated with increased risk<sup>3–5</sup>.

At a 2009 meeting, for example, Philippe Amouyel, chair of the EU Joint Programming Initiative on Neurodegenerative Diseases, compared data with Cardiff University geneticist Julie Williams, a long-time colleague. "We had found exactly the same genes," Amouyel recalls. "This was really important because it reinforces the fact that these genes were not just appearing through statistical bias."

The results have been further bolstered by validation in independent study groups, as well as by meta-analyses, which collectively examine multiple studies and assess their statistical power. "When people criticize GWAS, the best answer is that when we do a large, completely independent study, we get the same result," says Schellenberg.

The candidate genes also make biological sense, as most are involved with the inflammatory damage and metabolic disruptions that scientists have long associated with the disease (see 'Genetic risk factors for Alzheimer's disease'). "It's an assortment of genes that seem to be associated with lipid metabolism and immune response," says Richard Mayeux, co-director of Columbia University's Taub Institute for Research on Alzheimer's Disease and the Aging Brain in New York. "This was sort of predictable, but we didn't have the data to support it until now." Importantly, many of the genes also interact with the amyloid- $\beta$  pathway, which is still widely seen as the

initiating trigger for the disease (see 'Little proteins, big clues', page S12).

But these newly discovered genes do not resolve any debates about the origin of the disease — if anything, they potentially provide support for many different models of Alzheimer's pathogenesis. "Those who have been working on amyloid-independent pathways will say that genetics is proving it, while those working on amyloid will say, 'See, it's as we've said,'" says Christine Van Broeckhoven, a molecular geneticist affiliated with Belgium's University of Antwerp.

### DELIVERY TRUCK

Several of the candidate genes tie into multiple pathways, further complicating the picture. For example, clusterin (encoded by the *CLU* gene), which is one of the new risk factors most strongly associated with Alzheimer's disease, is thought to be involved in both amyloid- $\beta$  aggregation and clearance. It is also known as apolipoprotein-J, and is best known for helping ApoE facilitate cholesterol trafficking in the central nervous system. Another risk factor, complement receptor 1 (CR1), is an important component of the innate immune response against infection, but is also linked to the clearance of circulating amyloid- $\beta$ . But variants in genes such as *CLU* and *CR1* make relatively small contributions to the overall risk, increasing it by roughly 15%, so they have much less effect on the risk than *APOE*.

Exactly how ApoE might cause Alzheimer's disease is a matter of debate. As well as being the main transporter of cholesterol and other lipids and lipid-soluble molecules into the central nervous system, it is also thought to help remove amyloid- $\beta$  from the brain, although the mechanism is not yet clear. There are three major variants of the gene for ApoE. The protein produced by the high-risk *APOE4* variant is the least stable, significantly impairing the movement of cholesterol and amyloid- $\beta$  within the brain, whereas *APOE2* encodes a protein that is more abundant and actually confers protection against Alzheimer's disease relative to the common *APOE3* allele.

ApoE also modulates the inflammatory response to cellular damage in the brain, points out Thomas Montine, director of neuropathology at the University of Washington in Seattle. This reaction, mediated by the body's innate immune system, could be

**"We're now talking about seven strongly replicated genetic factors, all associated with lipid homeostasis."**

triggered by amyloid- $\beta$ -induced cell death, but it might also be a response to other neurological trauma, such as stroke. In either case, a prolonged inflammatory response can result in the gradual build-up

of toxic chemical by-products that further accelerate the death of neurons. Similar damage is seen in other neurodegenerative conditions, such as Parkinson's disease. "Almost all of the hypotheses are covered by *APOE4*," says Amouyel.

Several researchers are convinced that ApoE's role in cholesterol transport is the key to its importance in Alzheimer's disease. "The brain has 25% of the body's cholesterol content, even though it only makes up 2% of the body weight," says Judes Poirier, a neurobiologist at McGill University in Montreal, Canada. The brain's capacity to rewire itself, a property known as plasticity, depends on the ability to build and stabilize new synaptic connections. This in turn requires cholesterol, and mice that lack ApoE or express the *APOE4* variant exhibit dramatic problems in the repair of synaptic damage. "ApoE is your ultimate delivery truck when you need lipids to maintain or restore neural plasticity," Poirier says.

### MULTIPLE ROLES

This central role for ApoE is supported by evidence that variants in several other cholesterol-linked genes also increase the risk of Alzheimer's disease. One such gene is *PICALM*, which encodes a protein that assists ApoE in lipid traffic; another is *ABCA7*, which is also involved in cholesterol transport. "We're now talking about six or seven new, strongly replicated genetic factors, all associated with lipid homeostasis in the brain," says Poirier.

ApoE also seems to be a bridge between Alzheimer's disease and other physiological disorders. "The associations with cardiovascular disease and diabetes are strong — you very seldom find a study that doesn't show this association," says Mayeux. "The problem is, a stroke alone or the presence of diabetes alone doesn't cause the disease." But those who carry *APOE4* and have diabetes are twice as likely as non-diabetics with this variant to eventually develop Alzheimer's disease<sup>6</sup>.

Another piece of the *APOE4* puzzle is its link to a higher risk of heart attack and stroke. "That alone should be telling us that maybe its role here is actually lipid metabolism instead of some exotic amyloid- $\beta$ -interacting scheme," says Schellenberg. Accordingly, there is some evidence that taking statins, which lower cholesterol levels, may delay or prevent the onset of the cognitive decline associated with Alzheimer's disease, although clinical trials of statin use have yielded inconclusive results.

The available data fail to tie these various threads together satisfactorily, but several ambitious projects that are underway might help. For example, four of the largest Alzheimer's GWAS groups have joined forces, forming a mega-consortium known as the International Genomics of Alzheimer's Project. The project will draw on data from a total of 40,000 people with Alzheimer's disease and unaffected controls, and will



Genes regulating cholesterol are mutated in Alzheimer's; could statins be a treatment?

attempt a 'mega-meta-analysis', delving deeper in search of previously overlooked risk factors. "We're working with more than 10 million SNPs," says Amouyel. "That is very dense coverage of the genomic map."

The project also aims to identify which pathological features relate to specific genes. But differences in sample collection and storage across different groups are likely to complicate that goal. Van Broeckhoven points out that for many GWAS cohorts, researchers do not have access to a detailed medical history or post-mortem tissue collected using standardized autopsy protocols. This led to a lot of valuable disease data being lost before the study even began. "Knowing what we know today, we have to say that we have missed lots of opportunities in our sampling procedures," says Van Broeckhoven.

### EXPLORING EXOMES

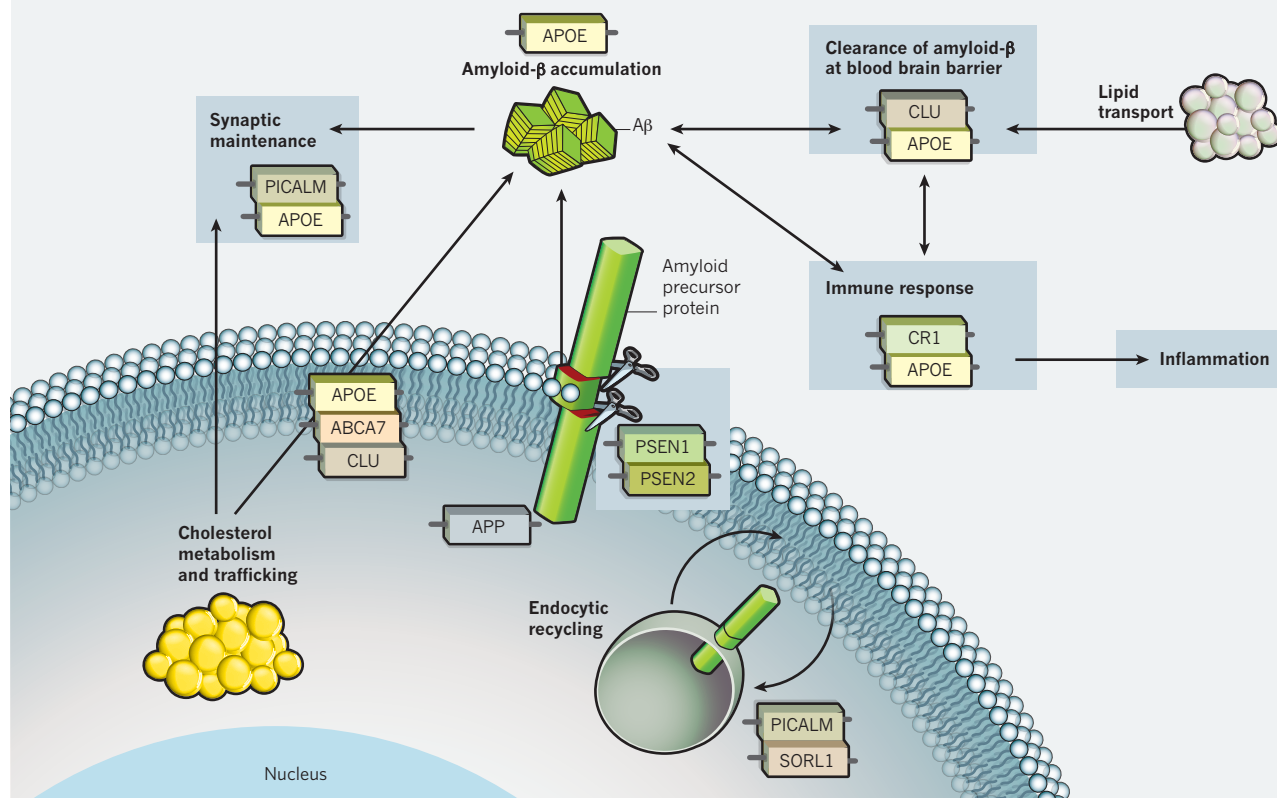
The GWAS studies are inherently limited by the distribution of known SNPs within the genome, leaving gaps that might conceal variants affecting the risk of disease. Because of the challenges of deriving statistically robust data for rare variants, these studies also typically ignore SNPs that are estimated to occur in less than 5% of the population.

However, the falling costs and increasing speed of DNA sequencing have made it easier



## GENETIC RISK FACTORS FOR ALZHEIMER'S DISEASE

Several genes implicated in Alzheimer's pathogenesis are involved in multiple cellular pathways, which illustrates the complexity of the disease.



for scientists to comb through entire genomes, and Schellenberg and colleagues are planning to use this approach to fill in the blanks. To save time and money, his team plans to focus initially on the exome — the subset of the genome that contains all the genes that are expressed — in the search for causal mutations. “I’d rather have 2,000 exomes sequenced than 100 genomes,” says Schellenberg, “because if you’re looking for something rare you need to have a big sample.”

Old-fashioned approaches to finding genes haven’t died out either, and several researchers are continuing to examine factors that were identified based on a hypothetical association with Alzheimer’s disease. For example, Mayeux’s group has identified several disease-associated SNPs within the *SORL1* gene, which encodes a protein that participates in the cellular uptake of APP. “There were a lot of doubters because it was a candidate gene, but it holds up in the latest GWAS,” says Mayeux. The role of *SORL1* is also supported by functional evidence: mice that produce lower levels of its protein accumulate more amyloid-β in the brain<sup>7</sup>.

Montine’s group identified another candidate while searching for physiological indicators in the blood or cerebrospinal fluid that might indicate the onset of Alzheimer’s

disease<sup>8</sup>. Brain-derived neurotrophic factor is linked to several other neurological conditions, and levels of this protein proved to be a powerful predictor of Alzheimer’s disease. However, there is no clear evidence of a causative role for variations in this gene. “We looked and couldn’t find an association, but we also haven’t sequenced the whole gene yet,” says Montine.

### A LIFETIME OF DAMAGE

A final component of risk is likely to emerge from the interface between genetic predisposition and physiological insults accumulated over the course of a lifetime. “In a disease that’s so strongly related to ageing, what we do and what we’ve been exposed to throughout our lives are likely to figure very importantly,” says Montine.

For example, diabetes and stroke can lead to the production of highly reactive compounds known as free radicals, which induce toxic chemical modifications in fats, proteins and nucleic acids. This sort of oxidative stress seems to be a general feature in the brains of people with Alzheimer’s disease, and could damage or kill neurons. “It’s a normal component of ageing, but there’s even more free-radical injury that occurs in people with Alzheimer’s,” says Montine. Mitochondria,

the energy centres of the cell, normally keep oxidative stress in check, and several studies are underway to assess whether mitochondrial DNA also contains risk factors for Alzheimer’s disease.

Attempts to understand the environmental aspect face the same problems that confront the geneticists: it is time consuming and expensive to acquire data, analyse it and then construct hypotheses that might prove meaningful for diagnosis, prognosis and treatment. “The genetics defines relevance but not mechanism,” says Montine, “and now it’s up to experimentalists to try to figure out how things work.” ■

**Michael Eisenstein** is a science writer based in Philadelphia, Pennsylvania.

1. Gatz, M. et al. *Arch. Gen. Psychiatry* **63**, 168–174 (2006).
2. Strittmatter, W. J. et al. *Proc. Natl Acad. Sci. USA* **90**, 1977–1981 (1993).
3. Bertram, L., Lill, C. M. & Tanzi, R. E. *Neuron* **68**, 270–281 (2010).
4. Hollingworth, P. et al. *Nature Genet.* **43**, 429–435 (2011).
5. Naj, A. C. et al. *Nature Genet.* **43**, 436–441 (2011).
6. Peila, R. et al. *Diabetes* **51**, 1256–1262 (2002).
7. Andersen, O. M. et al. *Proc. Natl Acad. Sci. USA* **102**, 13461–13466 (2005).
8. Zhang, J. et al. *Am. J. Clin. Pathol.* **129**, 526–529 (2008).





Rising life expectancy in developing countries such as China will bring with it an increase in the number of people with dementias.

# DEMENTIA

# A problem for our age

*As the number of Alzheimer's cases rises rapidly in an ageing global population, the need to understand this puzzling disease is growing.*

BY ALISON ABBOTT

The world is getting richer. But wealth brings its own burdens. Prosperous people live longer and old age carries a high risk of dementia — a condition that is so far neither preventable nor curable.

In 2000, for example, 4.5% of the population of the United States was older than 65 years, and there were 411,000 new cases of Alzheimer's disease. Ten years on, those numbers had risen to 5.1% of the US population and 454,000 cases, according to the Alzheimer's Association in the United States.

This same trend is happening across the world. In fact, when Alzheimer's disease is conflated with other dementias with similar clinical profiles, it covers an estimated 35.6 million people — around 0.5% of the global population. And these figures are about to get worse: the number of people with dementia is set to double in the next 20 years, according

to the *World Alzheimer Report 2010*, a global assessment of the economic impact of dementia.

Commissioned by Alzheimer's Disease International (ADI) — a federation of Alzheimer's associations around the world — the report gathered numbers on a range of Alzheimer's-like dementia. Dozens of teams are working to find ways to predict, prevent, diagnose and treat the condition, but so far their efforts have achieved only limited success. As a result, the economic costs of dementias will likely be crippling, the report says.

In 2010, the global economic impact of dementias was US\$604 billion. This figure dwarfs the costs of cancer or heart disease. Based on demographics, the ADI report foresees an 85% increase in cost by 2030, with developing countries bearing an increasing share of the economic burden.

"We are seeing a linear increase in prevalence in rich countries, but an exponential increase in low-income countries," says report co-author

Anders Wimo, an epidemiologist at the Karolinska Institute in Stockholm. "The need for solutions is urgent."

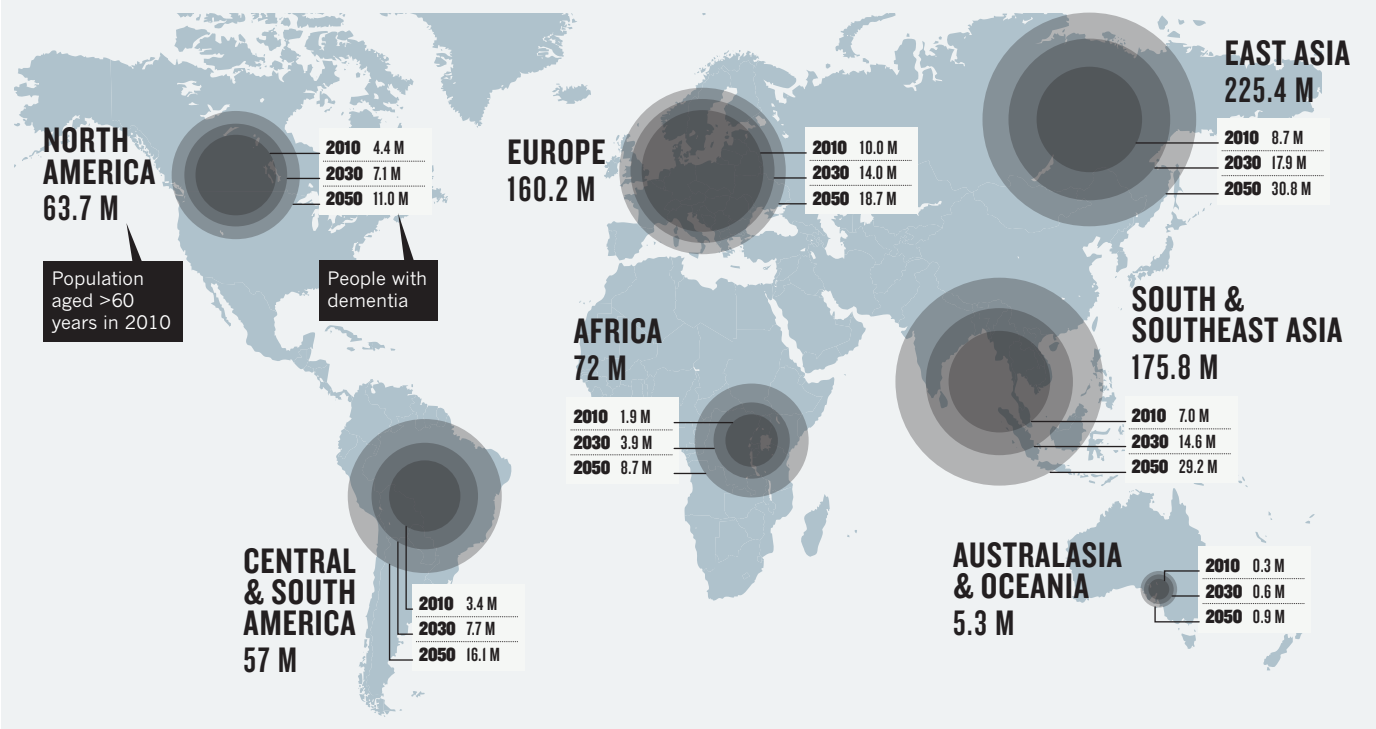
The ADI report used the best available data to determine the direct medical and social care costs, as well as the indirect costs, which mostly relate to family care and reduced productivity. Nearly 90% of the global costs in 2010, it says, are borne by rich countries — about 70% in Western Europe and North America — and less than 1% by low-income countries, where there is greater reliance on unpaid home care (see 'Global costs of dementia'). There is a fiftyfold difference in the cost of care per person between the richest countries and the poorest.

## AGEING IN ASIA

Just under half of people with dementia live in high-income countries, 39% live in middle-income countries, and only 14% live in low-income countries, the report says. But these

# ESTIMATED GROWTH OF DEMENTIA

The number of people with dementia will roughly double every 20 years, with the biggest increases in developing countries.



proportions are forecast to change dramatically in the coming decades, particularly in rapidly developing countries such as China and India, for two important reasons.

The first reason is demographic. In compiling the ADI report, Wimo and co-author Martin Prince of the Institute of Psychiatry at King's College London reviewed the available epidemiological studies. They found that the prevalence of dementias in people aged over 60 is fairly uniform across the world — between 5% and 7%.

As living standards increase in countries such as India and China, this will lead to increased life expectancy. Given that the biggest risk factor for dementia is age, a longer-living global population means there will be more people with dementia. The report predicts that the number of people with dementia will roughly double every 20 years, to 65.7 million in 2030 and 115.4 million in 2050 (see 'Estimated growth of dementia'). Most of this increase will be in developing countries.

Second, as wages rise, demand for more costly professional care will also increase — at least, that is what happened in wealthier countries where the Alzheimer's epidemic hit earlier. China has particular reason to worry: its one-child policy took effect in 1978, meaning that parents who reach old age in the next 20 years may not be able to rely on home care.

There are no comparable detailed global analyses for other chronic diseases. But *Dementia 2010*, a report commissioned by

the UK Alzheimer's Research Trust, estimated that the annual national cost of dementias was £23 billion (US\$38 billion), nearly twice that of cancer (£12 billion) and far more than the costs of heart disease (£8 billion) and stroke (£5 billion) (see 'Comparing costs').

The allocation of public research funds to these diseases does not reflect this hierarchy, however. In 2008, UK public spending on cancer research was 12 times higher than on dementia (see 'Comparing Investment'). In the United States, the National Institutes of Health spends 13 times more on cancer than on Alzheimer's-like dementias. "We can't fund all the good ideas we have in grant applications," says Neil Buckholtz, chief of dementia research at the US National Institute on Aging (NIA) in Bethesda, Maryland.

## TACKLING THE DISEASE

As the scale of the threat looms large, some countries are launching programmes to tackle dementia on several fronts. For example, in 2009, Germany opened the German Centre for Neurodegenerative Diseases (DZNE) in Bonn at a cost of €66 million (US\$95 million) per year. Developing treatment and preventive strategies will depend on clearly defining the disease and learning more about its clinical manifestations, says DZNE director Pierluigi Nicotera.

But these researchers will be aiming at a maddeningly elusive target. Fundamental questions about the disease — such as what its main cause is, and even what pathologies

define it — remain unanswered (see 'Common types of dementia'). The label 'Alzheimer's disease' was not widely used to describe dementia until 1976, when Robert Butler, the founding director of the NIA, coined the term, partly to make it easier to attract research funds to study the condition. At the time, the syndrome wherein some elderly people became forgetful and child-like was known as senile dementia. This was not viewed as a disease that might be prevented or cured, but as an intrinsic part of getting old.

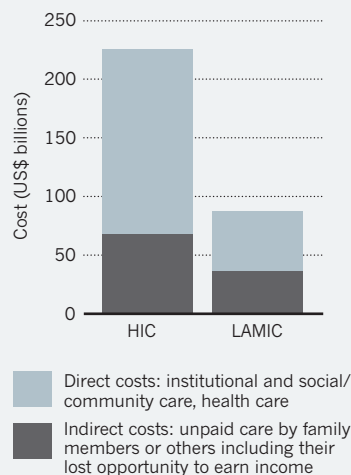
Alzheimer's disease is widely thought to be driven by amyloid pathology, in which peptides of amyloid- $\beta$  are generated in the brain and clump together into plaques. The plaques release toxic fragments of amyloid- $\beta$ , which wreak havoc by a mechanism that is not yet completely understood (see 'Little proteins, big clues', page S12).

Another form of dementia with similar symptoms is driven by vascular pathology. Leaking blood vessels deprive small areas of the brain of blood and oxygen, and these 'microstrokes' damage brain tissue and eventually result in cognitive defects. Scientists are still arguing about what proportions of dementias are driven by plaques and by vascular pathology. Post-mortem analyses of brains from people with dementia suggest that there is no simple answer: Alzheimer's-type pathology is more common, but it nearly always coexists with vascular pathology.

A 2011 investigation of more than 450 brains

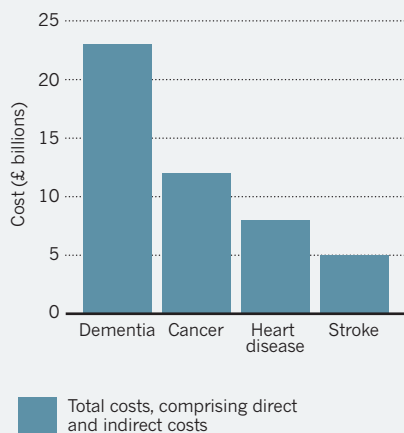
## GLOBAL COSTS OF DEMENTIA

There is a vast difference in the cost of care per person between high-income countries (HIC) and low- and middle-income countries (LAMIC).



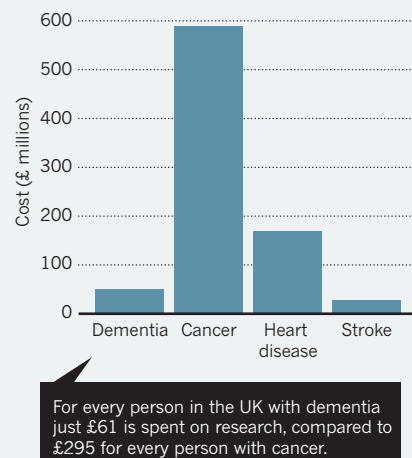
## COMPARING COSTS

In the United Kingdom, the economic impact of dementias dwarfs the costs of other diseases.



## COMPARING INVESTMENT

In the United Kingdom, annual government and charity spend on dementia research is 12 times lower than on cancer research.



from the UK Cognitive Function and Ageing Studies identified vascular damage in four-fifths of brains from individuals with dementia, and found plaques in nearly all of them (Wharton, S. B. *et al. J. Alzheimer's Disease*, in press). Scientists suspect that vascular pathology usually accelerates the damage driven by amyloid pathology. But the same study found that three-quarters of brains from individuals without dementia also had vascular pathology, and some of those from the oldest individuals showed a significant burden of plaques.

## AIMING AT AMYLOID

Amid this confusion, companies interested in developing therapies have primarily been targeting amyloid pathology, encouraged by the fact that the heritable, early onset form of Alzheimer's disease is mostly caused by mutations in the genes responsible for the production and metabolism of amyloid- $\beta$ . These familial cases account for fewer than 5% of total dementia (see 'Finding risk factors', page S20), but the companies hope that a significant proportion of later-onset dementia will be, one way or another, driven by amyloid- $\beta$ . "There is a level of wishful thinking in this," says

Nicotera. But so far none of the amyloid-based strategies has been successful (see 'A tangled web of targets', page S9). Yet drug developers have not given up on the concept. More reliable biomarkers of Alzheimer's disease are being developed (see 'Warning signs', page S5), potentially making it possible to carry out trials on patients before symptoms, and irreversible damage, set in.

Some scientists are also wondering whether it might be valuable to target vascular pathology as well. In fact, drugs such as statins, which lower cholesterol levels in the blood, and drugs to reduce blood pressure are now routinely given long term to patients at high risk of heart attack or stroke. If vascular pathology drives a significant proportion of dementias, those who have benefited from the long-term cardiovascular treatment introduced in the past two or three decades might be protected from dementias as well.

Few epidemiological surveys have so far backed this up, but the authors of the most rigorous survey to date, the Rotterdam Study, announced at the Alzheimer's Disease International conference in Toronto in March 2011 that they have observed a slowing in

the number of people being diagnosed with dementia.

Launched in 1990, the Rotterdam Study is considered to be a model for epidemiology trials. Intended to pinpoint the factors that contribute to various diseases, including dementia, in the elderly, it has recruited nearly 15,000 middle-aged individuals from a local population in three cohorts — in 1990, 2000 and 2006 — and is following their progress. Preliminary results have shown a small decrease in the age-specific incidence of dementias, and fewer plaques and less vascular damage among undiagnosed individuals, says epidemiologist Monique Breteler, head of the neurological and imaging part of the survey.

If dementias were ever to come under control, other medical problems of the elderly would become more prominent, notes Rudi Westendorp, who studies healthy ageing at Leiden University Medical Centre in the Netherlands. Because people with dementia are either less aware of pain or are unable to express their distress, "painful illnesses like herpes zoster [shingles] are probably being masked by dementia", he says. "Sight and hearing fail distressingly when we get old — we need to invest more heavily in research aimed at circumventing this, like developing neural implants to bypass damaged retinas."

Westendorp is an optimist who believes that solutions will be found to these problems, including the dementias, in the foreseeable future if countries invest in research now. Most of the problems that come with old age, he says, will have a medical solution — so living to a grand old age need not carry such a social and economic burden. ■

**Alison Abbott** is Nature's senior European correspondent.

## COMMON TYPES OF DEMENTIA

There is a great deal of overlap between the symptoms of various dementias.

Dementia type	Symptoms	Neuropathology	Proportion of dementia cases
Alzheimer's disease	Impaired memory, depression, poor judgement and confusion	Amyloid plaques and neurofibrillary tangles	50–80%
Vascular dementia	Similar to Alzheimer's disease, but memory less affected	Decreased blood flow to the brain owing to a series of small strokes	20–30%
Frontotemporal dementia	Changes in personality and mood, and difficulties with language	Damage limited to frontal and temporal lobes	5–10%
Dementia with Lewy bodies	Similar to Alzheimer's disease, also hallucinations, tremors	Cortical Lewy bodies (of the protein $\alpha$ -synuclein) inside neurons	<5%



# Warning signs

*The hunt is on for biomarkers that signal the descent into Alzheimer's disease. One initiative is leading the pack.*

BY RUTH WILLIAMS

Each week for the past six years, box after delivery box of blood, cerebrospinal fluid (CSF) and urine samples have arrived at a lab in the University of Pennsylvania in Philadelphia. Researchers there have documented, divided, labelled and stored the samples, row after row, in seven enormous freezers.

Some 14,000 samples have been divided into 160,000 tubes — and each one is precious. “We have back-up freezers and alarm systems in case of electrical failures,” says John Trojanowski, director of the Alzheimer's Disease Center at the University of Pennsylvania.

There's good reason for these precautions. The specimens, accompanied by detailed medical histories, cognitive and clinical measures, and high-resolution brain images, are among the “most highly annotated biological samples in the entire history of Alzheimer's disease research” — at least, that's the claim of the Alzheimer's Disease Neuroimaging Initiative (ADNI). Trojanowski is co-leader of ADNI's biomarker division.

At the moment, definitive diagnosis of Alzheimer's disease requires post-mortem analysis of the brain. While someone is still alive, the best bet is to assess their behaviour and memory, and rule out other disorders. Doctors are desperate for a marker that can reliably tell them who will get Alzheimer's disease, and what stage of the disease someone is going through.

A marker like that would, of course, be useful in the clinic, but it would also help researchers test drugs designed to slow the decline. The prevailing hypothesis in Alzheimer's disease is that deposition of the amyloid- $\beta$  protein leads to the formation of insoluble amyloid plaques between brain cells, and that these plaques are implicated in the dysfunction and death of brain cells (see ‘Little proteins, big clues’, page S12).

“Pharmaceutical companies were making drugs aimed at pulling the amyloid out, or reducing the amyloid, and they needed measures to monitor the effects of these treatments,” says Michael Weiner, professor of medicine, radiology and psychiatry at the

University of California, San Francisco, and ADNI's principal investigator. “Obviously, imaging and biomarkers were going to be important tools in all of this.”

## WORLDWIDE NETWORK

Launched in 2004, ADNI is one of the largest and longest-running studies of Alzheimer's disease<sup>1</sup>. Its goal is to find biological markers that can help determine how advanced someone's disease is and predict how well they will respond to treatment. The effort has already validated a few sensitive markers found by smaller studies.

This US\$160-million project is funded jointly by the US National Institutes of Health (NIH), 20 of the biggest pharmaceutical companies in the world, including Merck, AstraZeneca, Pfizer and GSK, and two non-profit partners, the Alzheimer's Association and the Alzheimer's Drug Discovery Foundation. “It is the largest public-private partnership that the NIH has,” says Weiner.

So far, ADNI has recruited 1,000 volunteers at 59 centres across the United States and Canada. Collaborative centres have also been set up in Europe, Japan, Australia and elsewhere. “What we are trying to do is establish a worldwide network of sites that are all using similar methods and sharing data,” says Weiner. “This makes it much easier to do international treatment trials and also allows us to look at differences between countries.”

ADNI is the best-funded effort in the hunt for Alzheimer's biomarkers, but it is by no means the only one (see ‘Finding risk factors’, page S20). Dozens of research teams are analysing brain images, DNA sequence variations and patterns in the expression of genes, proteins and immune molecules. In each case the aim is to identify measurable differences that either aid the diagnosis of Alzheimer's disease or reflect its progression.

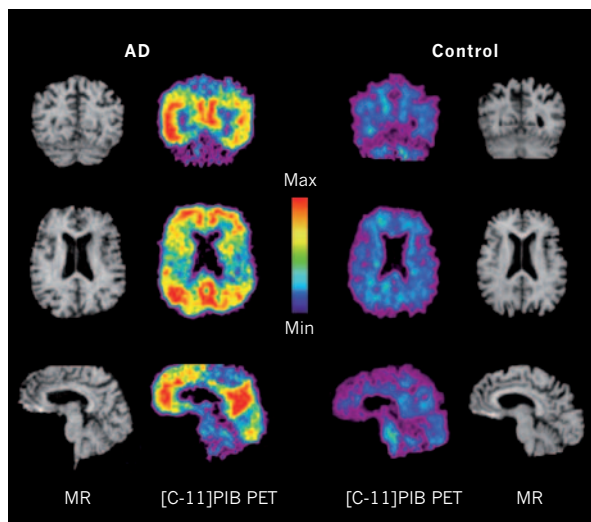
## A CLEAR PICTURE

Weiner says he wanted to do a multi-site study to compare different brain imaging techniques, such as magnetic resonance imaging (MRI) and positron emission tomography (PET), which could be used to detect changes in brain structure and metabolism associated with Alzheimer's disease. He approached several pharmaceutical companies, but the project was too expensive for any company to do it alone.

He then contacted Neil Buckholtz, chief of the Dementias of Aging Branch at the US National Institute on Aging (NIA). Buckholtz had been pondering a similar idea, so they began a series of discussions that led to the launch of ADNI a year later.

Of the 800 volunteers originally recruited, 200 had Alzheimer's disease, 400 had mild cognitive impairment (MCI) — a condition with high risk for progression to Alzheimer's disease — and 200 were healthy age-matched controls (including Weiner himself). After spending about a year standardizing operations and techniques, the team began using PET with a radioisotope of glucose known as FDG to measure brain metabolic activity, and using MRI to measure the volume of specific brain regions.

They also recorded levels in blood and CSF of various chemicals, including amyloid- $\beta$ , tau protein, sulphatides (components of nerve cell membranes), isoprostanes (markers of oxidative stress) and homocysteine (an amino acid), all of which had been shown to be altered in Alzheimer's disease. “ADNI's main goal has been to validate discoveries that were made in other smaller studies,” says Weiner, “and to show that these results really are replicable and clinically useful.”

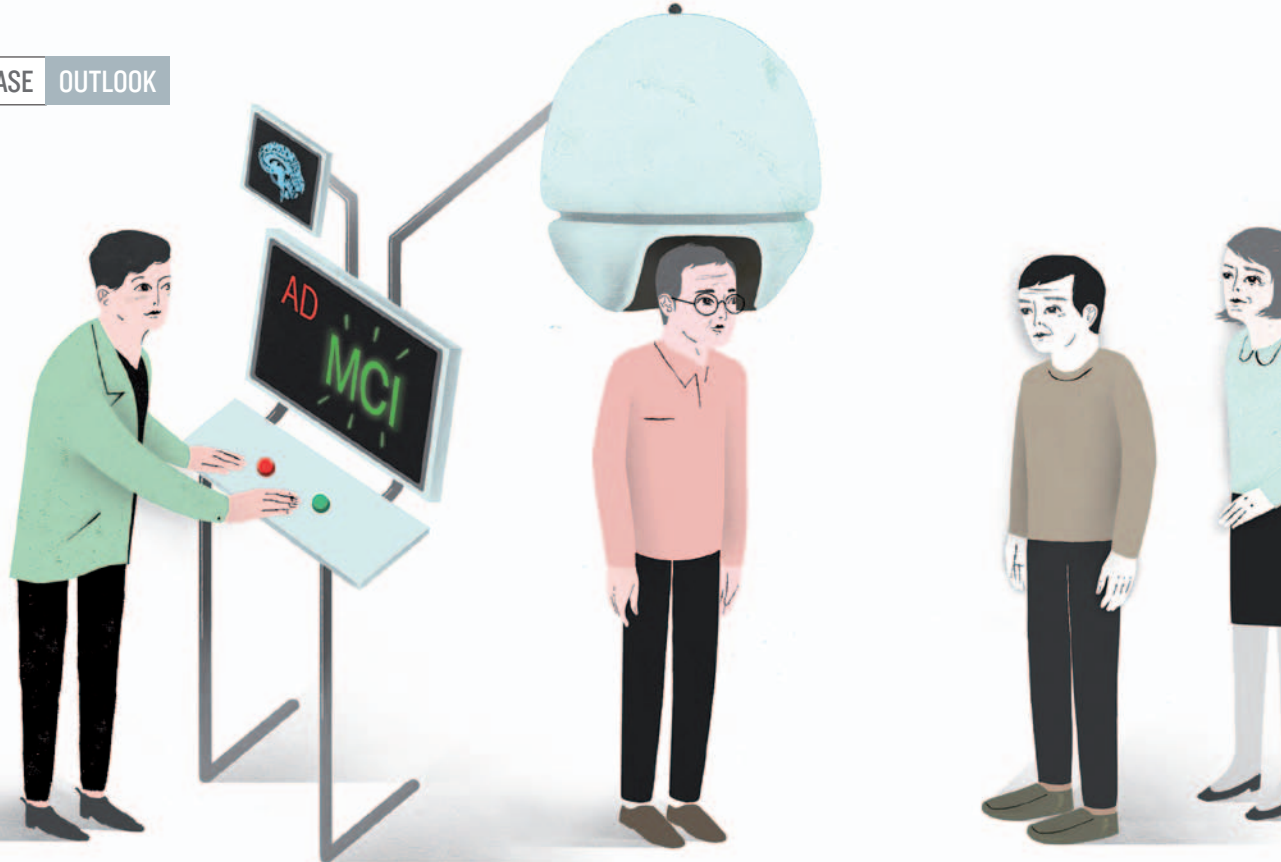


Pittsburgh compound B (PiB) lights up amyloid plaques in positron emission tomography (PET) images of the human brain.



ILLUSTRATION BY GRACIA LAM

UNIVERSITY OF PITTSBURGH, PENNSYLVANIA



### SENSITIVE MARKERS

As the data from these analyses emerged, some measures began to look quite promising while others fell by the wayside. The levels of sulphatides, isoprostanes and homocysteine in CSF, for example, turned out not to correlate to either Alzheimer's risk or disease progression. Another potential marker — the level of homocysteine in plasma — could help distinguish between MCI and healthy controls, but not between MCI and Alzheimer's disease.

Eventually, the researchers identified two sensitive biomarkers in CSF for detecting Alzheimer's disease, and for predicting the transition from MCI to Alzheimer's. One is the total level of tau protein; the other is the level of amyloid- $\beta$  — a 42-amino-acid peptide cleaved from amyloid precursor protein. The best CSF marker for indicating functional decline in healthy controls turned out to be P-tau, which is tau protein with additional phosphate groups<sup>2,3</sup>.

Imaging technologies are helping to identify changes in the brain that correlate with cognitive decline. MRI scans of people with advancing Alzheimer's disease reveal shrinkage of the temporal lobe and the hippocampus — the brain region used for storing memories and spatial navigation — and enlarged ventricles<sup>2</sup>, the brain cavities that contain CSF. The FDG-PET studies show that cognitive decline is most closely associated with reduced brain metabolic activity.

Shortly after the launch of ADNI, researchers at the University of Pittsburgh, Pennsylvania, developed a new form of PET. Using a radiolabelled compound called Pittsburgh

compound B (PiB), they generated scans that lit up amyloid plaques in the living human brain<sup>4</sup> (see image). ADNI quickly added this technique to its repertoire. In combination with CSF measurements, it confirmed that as levels of aggregated amyloid- $\beta$  in the brain increase, soluble amyloid- $\beta$  in the CSF diminishes. This not only established PiB-PET as a technique for detecting biomarkers but also further validated CSF amyloid- $\beta$  measures as reliable markers of brain pathology.

"I think we are still a little premature to say that these are validated biomarkers of prediction and progression, but it certainly is moving in that direction," says Ronald Petersen of the Mayo Clinic in Rochester, Minnesota, who heads the ADNI clinical core.

Despite Petersen's cautious endorsement, pharmaceutical companies are already using ADNI's measures in clinical trials. Meanwhile, ADNI continues to validate biomarkers in centres across the globe.

### EARLY STAGE

Kaj Blennow, professor of clinical neurochemistry at the University of Gothenburg, Sweden, and a member of the European ADNI, says that even if the biomarkers are robust enough to use, there are no reliable drugs to test them. "We need biomarkers in drug development," he says. "But at the same time, we need to have an approved drug that affects [amyloid- $\beta$ ] pathology or neurodegeneration so that we can use the drug to validate the biomarkers."

ADNI may have started out with the aim of validating and standardizing biomarkers, but its scope has grown well beyond. "ADNI has clearly shown that Alzheimer's

pathology in the brain exists in people long before they have dementia," says Weiner. The study has indicated that seemingly healthy people aged 70 years or above who have amyloid- $\beta$  in their brains might have a higher risk of developing dementia.

Indeed, new NIH guidelines for diagnosing Alzheimer's disease have expanded the definition of the disease to include MCI and a presymptomatic phase<sup>5</sup>. The presence of amyloid- $\beta$  at even this early stage could explain why trials of anti-amyloid- $\beta$  vaccines (see 'Chasing the dream', page S18) have been unsuccessful. Blennow says that the trials were carried out on patients with disease that was too advanced. "Perhaps the drug is not that effective when you have so much pathology, so you need to go earlier."

Whatever the reason, more long-term studies are needed that follow healthy people until a subset of them develops symptoms of Alzheimer's disease. With NIH funding secured for another six years, this is exactly what ADNI plans to do. The team has already recruited 200 new participants with early MCI. So far, "they are falling between the normal controls and the late MCI subjects", says Petersen. "It really is lining up somewhat as we expected and hoped."

### SIMPLE TEST

If pathology is present before subjects experience cognitive decline, then the logical next step would be the routine scanning of older adults to identify the telltale signs of the disease. But this is easier said than done. MRI is expensive and PET even more so and not readily available. "You only have PET



instruments in specialized large hospitals or research institutes,” says Blennow.

The lumbar punctures used to obtain CSF may be routine, but they are still much more invasive than drawing blood and carry a small risk of infection and damage to the spinal cord. “We cannot puncture healthy people or MCI patients,” says Christian Humpel, professor of psychiatry at Innsbruck University, Austria. “It’s not ethical.”

An ideal biomarker would show up in a simple blood test, and new markers that meet this criterion are regularly suggested. Candidates proposed in the past few years include clusterin, carbonyl proteins, angiotensin-converting enzyme, lipid peroxidation products and gene expression patterns. The ideal marker could be proposed next week, Humpel says, or it might not even exist. “We might have to use a combination of biomarkers.”

Humpel says he has unpublished evidence of two potential biomarkers — an immune molecule and a tumour-suppressor protein — found in blood monocytes, a type of immune cell. If other groups replicate his findings, these markers might end up in clinical screens, he says.

**“We need biomarkers in drug development and we need to have an approved drug to validate the biomarkers.”**

Like Humpel, Stanford University neurology professor Tony Wyss-Coray also thought that a combination of biomarkers might work best. In 2007, his team came up with a set of 18 plasma proteins that, measured together,

differentiate people with Alzheimer’s disease from healthy controls<sup>6</sup>. But even this approach did not lead to reproducible results.

“One reason you may not be able to reproduce a finding is because you use different tools,” says Wyss-Coray. His team used antibody arrays, which can be highly variable in the way they recognize and bind proteins, he explains. If all researchers used exactly the same array kit and plasma preparation techniques, they should get the same results, he says. Unfortunately, the kit his team used is no longer available.

#### ANTIBODY APPROACH

This lack of reproducibility has sounded the death knell for many promising biomarker studies, and it underscores the importance of ADNI’s efforts to standardize them. It also suggested to Thomas Kodadek, professor of chemistry and cancer biology at the Scripps Research Institute in Jupiter, Florida, that a different approach was required.

Instead of using an array of antibodies to look for proteins in the blood, Kodadek and his team decided to do the reverse: they are using an array of 15,000 synthetic proteins to look for antibodies in the blood. Antibodies are produced by the body’s immune system in response to foreign — or, in some cases, the body’s own — molecules, or antigens. “You are much better off trying to study antibodies rather than the antigens,” says Kodadek. “The antibodies shouldn’t be there at all in the absence of disease, but in the presence of disease they’re going to be amplified millionfold.”

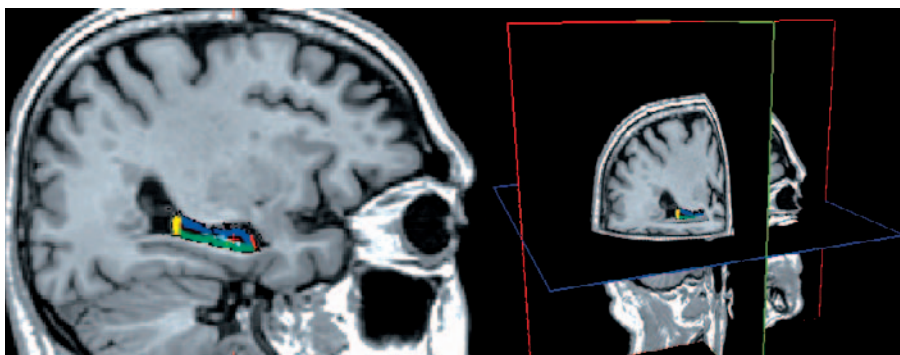
His approach assumes that the pathology of Alzheimer’s disease includes an immune response — an idea that is not generally shared among researchers. But his gamble seems to have paid off. His team has found two antibodies that are robustly expressed in 14 of 16 people with Alzheimer’s disease and just 2 of 16 control subjects<sup>7</sup>. Because the controls were age-matched, the two with high antibody levels might have preclinical disease, Kodadek says, in much the same way that amyloid plaques emerge well before cognitive symptoms. He has extended his study to about 200 people. “The results are holding up quite beautifully,” he says. “There are strong indications that our test is capable of picking up very early stage Alzheimer’s.”

Kodadek says he would like to test whether these antibodies are also amplified in blood samples from ADNI, and be able draw on all the associated imaging and other data. He’s not alone. ADNI is bombarded by requests from researchers who would like access to the samples, but cannot honour them all. After all, 160,000 tubes may sound like a lot, but they would quickly dwindle if every new candidate biomarker were tested. ■

**Ruth Williams** is a science writer based in Brooklyn, New York.

1. Alzheimer’s Disease Neuroimaging Initiative [www.adni-info.org](http://www.adni-info.org)
2. Trojanowski, J. Q. *et al. Alzheimer. Dement.* **6**, 230–238 (2010).
3. Shaw, L. M. *et al. Ann. Neurol.* **65**, 403–413 (2009).
4. Klunk, W. E. *et al. Ann. Neurol.* **55**, 306–319 (2004).
5. Jack, C. R. Jr *et al. Alzheimer. Dement.* **7**, 257–262 (2011).
6. Ray, S. *et al. Nature Med.* **13**, 1359–1362 (2007).
7. Reddy, M. M. *et al. Cell* **144**, 132–142 (2011).





Measuring the hippocampus, the brain structure most vulnerable to Alzheimer's, helps track the disease.

## PERSPECTIVE

# In search of biomarkers

New methods to follow changes in the brain or blood associated with Alzheimer's disease are critical for developing and testing drugs, says **Neil S. Buckholtz**.

This commentary starts from my frustration that no new drugs to combat Alzheimer's disease have been approved by the US Food and Drug Administration (FDA) since 2003. The few medications currently available address the symptoms of cognitive loss, but they do not delay or modify disease progression, and they work for only a limited time. In some people, they offer no relief at all.

It can take more than 10 years and almost US\$2 billion to bring a new drug into clinical use. Failure rates are high, especially for drugs that target the brain. And research into Alzheimer's disease is still stuck on a fundamental question. Although abnormal brain deposits of two proteins — amyloid- $\beta$  and tau — are hallmarks of the disease, we do not know if they are a cause or a by-product of the disorder.

Some Alzheimer's researchers believe that recently tested drugs might have failed because they were evaluated in people in the later stages of disease, when irreversible damage had already occurred and removing amyloid- $\beta$  was no longer beneficial. To modify disease progression, therapies need to be applied earlier and be targeted at the disease mechanisms occurring in the brain at that stage. We therefore need better biomarkers to provide insight into disease progression and to help target drugs to the right pathological processes at the right time.

We are making strides. This year, the National Institute on Aging (NIA), which is part of the US National Institutes of Health (NIH), and the Alzheimer's Association jointly introduced a different way of thinking about Alzheimer's disease. These new diagnostic guidelines — the first in 27

years — should not only guide research but also enhance the development and testing of interventions. They present three stages of Alzheimer's disease: preclinical<sup>1</sup>, mild cognitive impairment<sup>2</sup> (MCI), and dementia<sup>3</sup>.

The guidelines also address whether changes in the brain, blood and cerebrospinal fluid (CSF) are associated with Alzheimer's disease. Researchers often use such biomarkers to detect the onset of disease and track its progression, but they cannot become

routine in clinical diagnosis without further testing and validation. I hope that testing and validation of biomarkers under the revised guidelines will result in improved diagnosis and treatment of Alzheimer's disease. Importantly, the guidelines will inform discussions about clinical trials for disease prevention (see 'Prevention is better than cure', page S15).

Another effort relating to biomarkers is the Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI is a public-private partnership comprising scientists from academia and from pharmaceutical and diagnostics companies. ADNI is identifying the best methods — such as clinical and neuro-psychological tests, magnetic resonance imaging (MRI, see top image), positron emission tomography (PET), genetic screening, and blood and CSF markers — to evaluate the progression from normal cognitive ageing to MCI, from MCI to mild Alzheimer's dementia, and from

mild to severe dementia, and to use these markers for early diagnosis. These biomarkers are being incorporated into clinical trials to assess the ability of new drugs to modify disease progression. The ADNI data are freely available to the international scientific community.

However, there is more work to be done. In preclinical drug discovery and development, biomarkers must provide a bridge to clinical studies to allow accurate prediction of treatment response in humans, as animal models are currently poor predictors of outcomes in human clinical trials. In the clinical arena, the major problem arises in the transition from phase II to phase III clinical trials — many failures in phase III had a positive signal in phase II. New biomarkers are needed to ensure that a drug candidate is engaging the correct target and that the proper dose has been selected at phase II.

The NIA has established a variety of funding mechanisms to meet these challenges. Many of the efforts involve biomarkers: preclinical translational biomarkers that help predict clinical therapeutic potential, and clinical biomarkers that indicate target engagement and dose selection. In the best of all worlds, the same neuroimaging (MRI, PET) and fluid (blood, CSF) biomarkers could be used for both preclinical and clinical use.

I believe that cooperation between public and private organizations in the preclinical arena — equivalent to ADNI's clinical role — will ease the transition from preclinical to clinical work. This idea is fairly radical for a preclinical setting and would require new ways of thinking from the private sector; in particular, companies would need to share information with one another, as well as with academia. This kind of sharing, perhaps brokered by the NIH and the Foundation for the NIH (which manages the private partners in ADNI), could advance the use of biomarkers with the goal of developing a standard set of criteria for assessing preclinical drug efficacy and aiding clinical decision-making. The greater use of biomarkers made possible by this strategy should help to advance the discovery of new drugs — and benefit the patients and families who desperately need them.■

**Neil S. Buckholtz** is chief of the Dementias of Aging Branch in the Division of Neuroscience at the National Institute on Aging in Bethesda, Maryland.  
e-mail: buckholn@nia.nih.gov

1. Sperling, R. A. et al. *Alzheimer's & Dement.* **7**, 280–292 (2011).
2. Albert, M. S. et al. *Alzheimer's & Dement.* **7**, 270–279 (2011).
3. McKhann, G. M. et al. *Alzheimer's & Dement.* **7**, 263–269 (2011).



## DRUGS

# A tangled web of targets

*Drugs in development for Alzheimer's disease take aim at a variety of neural mechanisms. But despite a wealth of possibilities, there have been few successes.*

BY LAUREN GRAVITZ

For research into Alzheimer's disease, 2003 was a good year. The US Food and Drug Administration (FDA) had just approved memantine, the first in a class of drugs that reduces abnormal brain activity. Scientists had identified several potential targets, various academics and companies were developing therapies based on each, and the field seemed to be moving in the right direction.

Memantine is one of several drugs on the market in Europe and the United States that can slow the mental and physical decline of patients already in the throes of Alzheimer's disease. These drugs boost the activity of healthy neurons in the brain, masking the progression of dementia for a limited time (memantine, for example, seems to be effective for at least six months). But none of them can stop Alzheimer's disease in its tracks. So researchers began to shift their emphasis from treating symptoms to attacking the underlying cause of disease.

Eight years on, multiple therapies are in late-stage testing (see 'Selected drugs in clinical trials 2011'), including four that have the

potential to modify the biological roots of Alzheimer's disease<sup>1</sup>. And yet, despite these seemingly imminent improvements in Alzheimer's therapeutics, a vague pall of scepticism hangs over the field. There is no clear evidence that these approaches will work, and many indications that they may not.

"Progress in the basic science of disease has been so substantial for the last few decades that many of us were quite optimistic," says Paul Aisen, director of the Alzheimer's Disease Cooperative Study and a researcher at the University of California, San Diego. But a sense of stasis has now set in. "We haven't had a new drug since 2003," says Aisen, "and the result of every major trial that's reported since then has been very disappointing."

## A TWISTED TALE

A big part of the problem is that researchers don't know enough about the biology of Alzheimer's disease to identify the right targets. The disease is the result of a long chain of events, but some of the links in that chain are still a mystery — nobody is certain which link to cut to stop disease progression.

In a field with limited funding, the multitude of theories and possible targets has made for a difficult, albeit stimulating, challenge. "The therapeutic landscape for Alzheimer's disease is wide open — and it's wide open because we don't have a good definition of the disease, we don't have validated drug targets, and we have too many unvalidated ones," says Lon Schneider, a gerontologist and neurologist at the University of Southern California in Los Angeles.

But despite the wide variety of potential approaches, three of the four drugs in phase III trials share one main target: an improperly folded peptide called amyloid- $\beta$ . In people with Alzheimer's disease, this protein fragment is sequestered into hard plaques nestled between neurons in the brain. Although few researchers doubt that amyloid- $\beta$  is at least partly to blame for the disease (see 'Little proteins, big clues', page S12), many are beginning to wonder whether it is the right molecule to target.

Amyloid plaques are one of the hallmarks of Alzheimer's disease. Indeed, imaging studies have shown that plaques can start to accumulate 10–15 years before symptoms

emerge, prompting researchers to suggest that amyloid- $\beta$  may be a good target for prevention (see 'Prevention is better than cure', page S15). Eliminating amyloid- $\beta$  might not halt the disease, however. By the time Alzheimer's becomes symptomatic, attacking amyloid- $\beta$  could have no perceptible effects.

So maybe a different type of drug is needed to halt or reverse cognitive decline. It might be better, some researchers suggest, to target another characteristic of the disease: the twisted clumps of fibrous protein inside neurons called neurofibrillary tangles. These are caused by the accumulation of a toxic form of the tau protein and correlate closely to the timing of symptom onset.

Other researchers champion wholly different approaches, ranging from brain surgery to repurposing drugs approved for a host of conditions including diabetes and arthritis. "This is a messy illness, and there are many, many ways of potentially cleaning up the mess," says Schneider. "That's what's so frustrating."

## PREVENTING CLEAVAGE

The evidence pointing to amyloid- $\beta$  as a cause of Alzheimer's disease seems overwhelming. Genetic studies reveal abnormal amyloid- $\beta$  production in familial Alzheimer's disease, and cell-culture and animal studies implicate the misfolded protein in everything from neuronal death to behavioural and memory problems. For nearly two decades, most of the therapeutic research has focused on finding ways to reduce amyloid- $\beta$  production and dissolve amyloid plaques in the brain.

But targeting amyloid- $\beta$  is far from simple. "There's a slew of uncertainty about where in the disease course one would have to intervene" to target amyloid- $\beta$ , says Jeffrey Cummings, director of the Cleveland Clinic's Lou Ruvo Center for Brain Health in Las Vegas, Nevada. "Its biology is very complex, the pathways for amyloid- $\beta$  metabolism are multiple, and it may prove to be very difficult to work with."

Two different enzymes —  $\gamma$ - and  $\beta$ -secretase — cleave the amyloid precursor protein (APP) in two different spots, separating the short amyloid- $\beta$  peptide from its progenitor. These peptides aggregate into small, stable clusters called oligomers, which then clump together to form larger plaques. Every step of the process, from the first snip to the final plaque, presents an opportunity to arrest the disease.

One approach involves  $\gamma$ - and  $\beta$ -secretase inhibitors. If  $\gamma$ - and  $\beta$ -secretase can be prevented from cleaving APP in the first place, there will be no amyloid- $\beta$ . But targeting these enzymes has proven tricky, partly because  $\gamma$ -secretase is not specific to APP but also cleaves other proteins — including the vital protein Notch. One of the greatest disappointments in Alzheimer's therapeutics so far came last year when Eli Lilly, of Indianapolis, Indiana, abruptly halted a phase III trial of its  $\gamma$ -secretase inhibitor, semagacestat,



when an interim analysis revealed that the drug actually accelerated the progression of disease rather than slowing it down.

The reasons for this failure are still being investigated. "One possibility is that maybe anything you do that manipulates amyloid- $\beta$  is bad for the brain," says Eric Seimers, the Eli

*"Amyloid- $\beta$  biology is very complex and it may prove to be very difficult to work with."*

Lilly senior medical director who oversaw the trial. "A more likely possibility, though, is that the worsening is not because we reduced amyloid- $\beta$ , but because  $\gamma$ -secretase does something else that the brain needs." Several of the trial subjects also developed skin cancer, perhaps because of the drug's effects on Notch.

So what about  $\beta$ -secretase? It is more specific to APP than  $\gamma$ -secretase, and pharmaceutical companies are in dogged pursuit of drugs that inhibit it. But the enzyme's shape has turned out to be problematic. Researchers have had a difficult time creating a molecule that is large enough to inhibit the enzyme's active binding site but small enough to pass through the blood-brain barrier so it can be taken orally. Despite some early setbacks, many companies — Eli Lilly included — are continuing to target  $\beta$ -secretase, and a few compounds are in early stage trials.

## AIMING AT AMYLOID

All eyes, however, are trained on a passive immune therapy that leaves both secretase enzymes alone and goes after amyloid- $\beta$  directly. Two candidates, Eli Lilly's solanezumab and Janssen and Pfizer's bapineuzumab (originally developed by the Dublin-based company Elan), are monoclonal antibodies that work with the immune system, binding to amyloid- $\beta$  and helping to clear accumulated amyloid- $\beta$  peptides in the brain. Both are being tested in phase III trials on thousands of participants with mild-to-moderate Alzheimer's disease (see 'Chasing the dream', page S18).

"Probably every big company and even a number of smaller companies have products that will eliminate amyloid in an amyloid-producing mouse," says William Thies, chief medical and scientific officer at the Alzheimer's Association, a nonprofit organization in Chicago, Illinois, dedicated to patient care and research funding. "But they're not going to move them on to a phase III study until they see the results of the ongoing trials. They would like to be convinced that the amyloid hypothesis is correct."

There are some hints that it might not be — or at least that targeting amyloid- $\beta$  will not work once symptoms are apparent. The 18-month phase II trial of bapineuzumab left many researchers feeling sceptical. Although imaging studies showed that the antibody decreased amyloid plaques in the brain<sup>2</sup>, it seemed to have little, if any, effect on cognition. With vaccine studies yielding similar results, many are growing uneasy with the approach, and suggest that it might work only as a preventive measure and should be tested in people without symptoms<sup>3</sup>. Attacking amyloid plaques in symptomatic patients may be like cleaning up the mess inside a house after a flood: the structure remains, but all the personal effects are long gone.

Others say that plaques could be the body's way of sequestering the toxic amyloid- $\beta$  oligomers. Elan, which led the field in immune approaches, has a candidate called scyllo-inositol. "It binds to amyloid- $\beta$  at some intermediate structure, blocking its ability to form plaques and also blocking its ability to cause toxicity to neurons," says Dale Schenk, Elan's chief scientific officer.

But some worry that researchers are spending too much time and resources on something that might never pan out. "I think amyloid- $\beta$  is proving to be a very intractable target," says Cummings. "The great danger to the field is that if bapineuzumab fails, some pharmaceutical companies will decide that Alzheimer's disease is too tough a target to yield stockholder value and will redirect their resources toward more tractable diseases."

## TAUIST PHILOSOPHY

With the amyloid- $\beta$  issue still unresolved, more researchers are looking to the second major target: tau protein.

Tau protein normally stabilizes structural elements, called microtubules, in healthy neurons. In Alzheimer's disease and other 'tauopathies', however, tau acquires too many phosphate groups and becomes dysfunctional. It aggregates inside neurons, the microtubules collapse, and the resulting neurofibrillary tangles block neuronal signalling.

Neither amyloid plaques nor tau tangles are solely responsible for causing Alzheimer's disease, but of the two, tangles show a better correlation with clinical symptoms, says Peter Davies, director of Alzheimer's research at the Feinstein Institute for Medical Research in Manhasset, New York. "You can have a lot of amyloid in your



## SELECTED DRUGS IN CLINICAL TRIALS 2011

Drug	Trial status	Mode of action	Developer
Bapineuzumab	Phase III, ongoing	Humanized monoclonal antibody to amyloid- $\beta$ ; targets the peptide's N-terminus	Pfizer/Janssen
Solanezumab	Phase III, ongoing	Humanized monoclonal antibody to amyloid- $\beta$ ; targets the centre of the peptide	Eli Lilly
Intravenous immunoglobulin (IVIg)	Phase III, ongoing	Isolated from pooled human blood, believed to have anti-amyloid- $\beta$ and anti-inflammatory properties	Baxter
Latrepidine (Dimebon)	Phase III, ongoing	Thought to stabilize mitochondria, thereby protecting neurons and preventing them from malfunctioning	Pfizer/Medivation
Scyllo-inositol / ELND 005	Phase II completed, Phase III in planning	Prevents or inhibits amyloid- $\beta$ aggregation	Elan
Methylthioninium chloride (Rember)	Phase II completed, Phase III in planning	Unclear; thought to inhibit tau aggregation, but may be acting as an anti-amyloid- $\beta$ disaggregator	TauRx Pharmaceuticals
CERE-110	Phase II, ongoing	Adenovirus-aided delivery of a nerve growth factor gene that helps protect neurons; delivered via surgery	Ceregene
PBT2	Phase IIb in planning	Metal chelator, small molecule that inhibits tau hyperphosphorylation and amyloid- $\beta$ aggregation	Prana Biotechnology
Davenutide/AL-108	Phase II completed	Microtubule stabilizer, preventing tau hyperphosphorylation and tangle formation	Allon
BMS-708163	Phase II, ongoing	Inhibits formation of $\gamma$ -secretase, thereby inhibiting formation of amyloid- $\beta$	Bristol-Myers Squibb
PF-04494700/TTP488	Phase II, ongoing	RAGE inhibitor, modulates glial activity and reduces amyloid- $\beta$ plaque formation	Pfizer
Tideglusib/NP-12 (Nypta)	Phase II, ongoing	GSK-3 inhibitor, preventing tau hyperphosphorylation	Noscira

brain and be absolutely fine," he says. "If you have a lot of tau pathology, you're never fine."

Self-dubbed 'tauists', who believe that tau protein is the key to Alzheimer's disease, are studying whether interfering with the extra phosphate groups<sup>4</sup> or the enzyme that attaches them could slow or even reverse the symptoms of disease. "Until you can undo tau pathology and show that it undoes symptoms, you won't know for sure," Davies says.

Tau research has progressed more slowly than work on amyloid- $\beta$ , partly because of scant funding and the overwhelming interest in amyloid- $\beta$ , and partly because of tau's essential role in maintaining healthy cells. But a few groups have persisted, and at least one drug candidate has made it to phase II trials. In April, Madrid-based biopharmaceutical company Noscira began European efficacy trials on a compound that inhibits GSK-3, the enzyme that adds phosphate groups to tau. This is actually the second GSK-3 inhibitor to go to human testing — lithium inhibits the same enzyme, but small trials of lithium were inconclusive.

One of the most hyped therapies in the tau class is a repurposed drug. In 2008, researchers from TauRx Pharmaceuticals in Singapore

made an announcement that sent small tremors of excitement through the field. They tested a modified version of methylene blue — an outdated treatment for malaria, urinary tract infections and bipolar disorder — in a dosing and efficacy trial of 321 people with mild-to-moderate Alzheimer's disease. After 84 weeks, the cognitive decline of those on the drug appeared to be 81% slower than those taking a placebo.

At the time, TauRx scientists claimed that their drug, which they call Rember, worked by preventing tau aggregation. But the company never published the data — the compound is old, with a proven safety record, so they had no regulatory incentive to do so. Their reported results still remain a mystery. An independent team of researchers found that, in animal models at least, methylene blue clears amyloid- $\beta$  but has no effect on tau whatsoever<sup>5</sup>. Despite these contradictory interpretations, the company is seeking patents in both Europe and the United States, and announced in December that it plans to press ahead with a large-scale phase III trial.

## SLOWING THE DECLINE

Alzheimer's disease is one of dying neurons — of incomplete circuits, of neural impulses left

unfinished, of thoughts, memories and ideas that are lost with the dying nerves. Thus, other than those therapies that target amyloid- $\beta$  or tau protein, most of the drugs in the pipeline aim to protect some of these neurons and to slow the course of disease, not reverse it.

Among the more radical approaches is CERE-110 from Ceregene in San Diego. The product delivers a gene for nerve growth factor (NGF) — a large protein that helps nurture neurons — where it is most needed. Because NGF cannot pass through the blood-brain barrier, and because the gene must be incorporated by the specific subset of neurons most affected by the disease, precise delivery requires brain surgery.

The seriousness of brain surgery cannot be overlooked, but this particular procedure takes just a few hours and no safety problems have been reported from more than a hundred operations done so far, says Mark Tuszynski of the University of California, San Diego, the researcher who devised the approach. "It won't be a cure," he says, "but the hope is that it can meaningfully enhance neurons enough to slow decline and have a useful impact on quality of life."

Dimebon, which has been used as an anti-histamine in Russia since 1983, has also shown an ability to protect neurons. In 2008, a small phase II trial of dimebon in 155 people with mild-to-moderate Alzheimer's disease yielded surprisingly good results: those taking the drug appeared to improve in cognition and daily function for up to 12 months.

But the excitement over the drug's potential abated when a larger trial failed to elicit the same results. In 2010, the first large phase III trial of dimebon showed no significant difference between the test and control groups. A second phase III trial is underway. "Something like dimebon comes along, with no rational reason about why it works the way it's supposed to work, and people go gaga and suspend their ordinary scientific scepticism," says Schneider at the University of Southern California.

Alzheimer's disease hides its secrets well. So although researchers may disagree on the best approach for halting it, most agree that the current range of targets provides a good starting point. "There are lots of things in the pipeline — lots of different possibilities. And that's what we need at this time," says Thies of the Alzheimer's Association. "The reality is that nobody knows which approach will be best." ■

**Lauren Gravitz** is a science writer based in Los Angeles, California.

1. Mangialasche, F. *et al. Lancet Neurol.* **9**, 702–716 (2010).
2. Rinne, J. O. *et al. Lancet Neurol.* **9**, 363–372 (2010).
3. Golde, T. E. *et al. Neuron* **69**, 203–213 (2011).
4. Mazanetz, M. P. & Fischer, P. M. *Nature Rev. Drug Discov.* **6**, 464–479 (2007).
5. Medina, D. X., Caccamo, A. & Oddo, S. *Brain Pathol.* **21**, 140–149 (2011).

# Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF

Cameron D. Mackereth<sup>1,2,3</sup>, Tobias Madl<sup>1,4</sup>, Sophie Bonnal<sup>5</sup>, Bernd Simon<sup>3</sup>, Katia Zanier<sup>3</sup>, Alexander Gasch<sup>3</sup>, Vladimir Rybin<sup>3</sup>, Juan Valcárcel<sup>5,6</sup> & Michael Sattler<sup>1,3,4</sup>

Many cellular functions involve multi-domain proteins, which are composed of structurally independent modules connected by flexible linkers. Although it is often well understood how a given domain recognizes a cognate oligonucleotide or peptide motif, the dynamic interaction of multiple domains in the recognition of these ligands remains to be characterized. Here we have studied the molecular mechanisms of the recognition of the 3'-splice-site-associated polypyrimidine tract RNA by the large subunit of the human U2 snRNP auxiliary factor (U2AF65)<sup>1–3</sup> as a key early step in pre-mRNA splicing<sup>4</sup>. We show that the tandem RNA recognition motif domains of U2AF65 adopt two remarkably distinct domain arrangements in the absence or presence of a strong (that is, high affinity) polypyrimidine tract. Recognition of sequence variations in the polypyrimidine tract RNA involves a population shift between these closed and open conformations. The equilibrium between the two conformations functions as a molecular rheostat that quantitatively correlates the natural variations in polypyrimidine tract nucleotide composition, length and functional strength to the efficiency to recruit U2 snRNP to the intron during spliceosome assembly<sup>1,5–8</sup>. Mutations that shift the conformational equilibrium without directly affecting RNA binding modulate splicing activity accordingly. Similar mechanisms of cooperative multi-domain conformational selection may operate more generally in the recognition of degenerate nucleotide or amino acid motifs by multi-domain proteins<sup>9,10</sup>.

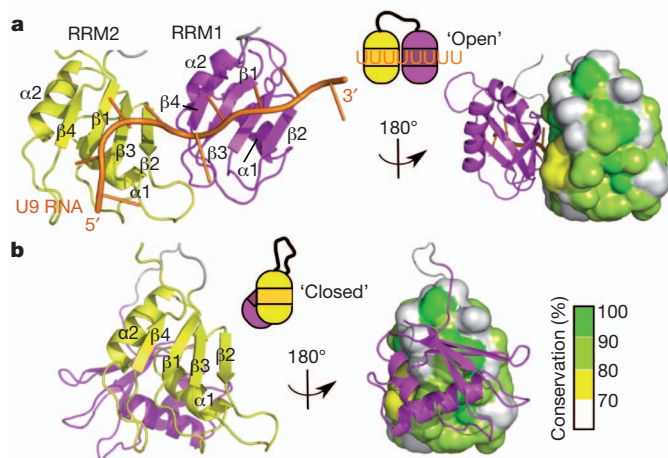
The essential multi-domain splicing factor U2AF65 has a crucial role in the assembly of splicing complexes<sup>4</sup>. A polypyrimidine (Py) tract RNA sequence at the 3' end of introns is recognized by the tandem RNA recognition motif (RRM) domains (RRM1–RRM2) of U2AF65 (refs 11, 12). However, there is significant diversity in the nucleotide composition, length and functional strength of the Py tract sequence, reflecting the dynamic range of splice site acceptor site usage in events such as alternative splicing. The mechanisms by which the Py tract sequence variations found in human U2 introns<sup>1,5–8</sup> are recognized by U2AF65 and how the 'strength' of a given Py tract is coupled to the efficiency of spliceosome assembly are not understood. Using a novel protocol for structural analysis of multi-domain proteins and protein complexes in solution<sup>13</sup> (Supplementary Text and Methods), we studied the minimal region in U2AF (U2AF65 RRM1–RRM2, residues 148–342; Supplementary Fig. 1a) that mediates binding to the Py tract RNA and recapitulates the key features of Py tract recognition by U2AF (Supplementary Text and Supplementary Figs 2–4).

We found that the U2AF65 RRM1–RRM2 tandem domains can populate two distinct three-dimensional arrangements correlated to the presence or absence of a high-affinity RNA ligand (Fig. 1a, b and Supplementary Tables 1 and 2). In the 'open' conformation of the RRM1–RRM2 tandem domains, as observed when bound to U9 RNA (Fig. 1a), a parallel arrangement of the two  $\beta$ -sheets forms an extended basic RNA-binding surface (Supplementary Fig. 5). The protein–protein

interface between the two RRM domains involves residues from  $\alpha 2$  to  $\beta 4$  in RRM1 and  $\alpha 1$ ,  $\beta 2$  and the  $\beta 2$ – $\beta 3$  linker in RRM2, stabilized mainly through electrostatic complementarity. The RRM1–RRM2–U9 model also incorporates atomic details of protein–RNA contacts for the individual RRM domains seen in the previous crystal structure<sup>12</sup> (Supplementary Fig. 6). However, the nuclear magnetic resonance (NMR) data are inconsistent with the overall arrangement of the tandem RRM domains in the crystal (Supplementary Fig. 7), indicating that the relative domain orientation was influenced by crystal packing forces and/or deletion of the linker (which is conserved in length, Supplementary Fig. 8).

In the second, 'closed' conformation, observed in the absence of ligand (see Supplementary Text for details on structure calculation), the RNA-binding surface of RRM1 ( $\beta 2$ ) is partially occluded by an interaction with helices  $\alpha 1$  and  $\alpha 2$  of RRM2 (Fig. 1b). The protein–protein interface between RRM1 and RRM2 in this 'closed' conformation agrees well with residues identified based on chemical shift differences between RRM1–RRM2 and the isolated domains (Supplementary Fig. 9c, d) and is further supported by an excluded solvent-accessible area (derived from solvent paramagnetic relaxation enhancement (PRE) data; not shown). As in the RNA-bound form, the domain interface comprises mainly electrostatic interactions involving conserved residues (Supplementary Fig. 8).

One set of measurements required for model generation involves long-range distance restraints derived from PRE. PREs are obtained by spin labelling various residues in RRM1–RRM2 and are detected as



**Figure 1 | Structure of the tandem RRM domains of U2AF65 free and when bound to a high-affinity Py tract.** **a, b,** Cartoon and ribbon representation of the lowest energy solution structure models calculated for the (a) RNA-bound or open form of RRM1–RRM2 with a U9 Py tract RNA (orange), and (b) the unbound or closed form of RRM1–RRM2. The conserved surface of RRM2 is exposed in the open conformation (a, right) but is occluded by RRM1 (shown as magenta ribbon) in the free protein or in the presence of weak Py tracts (b, right).

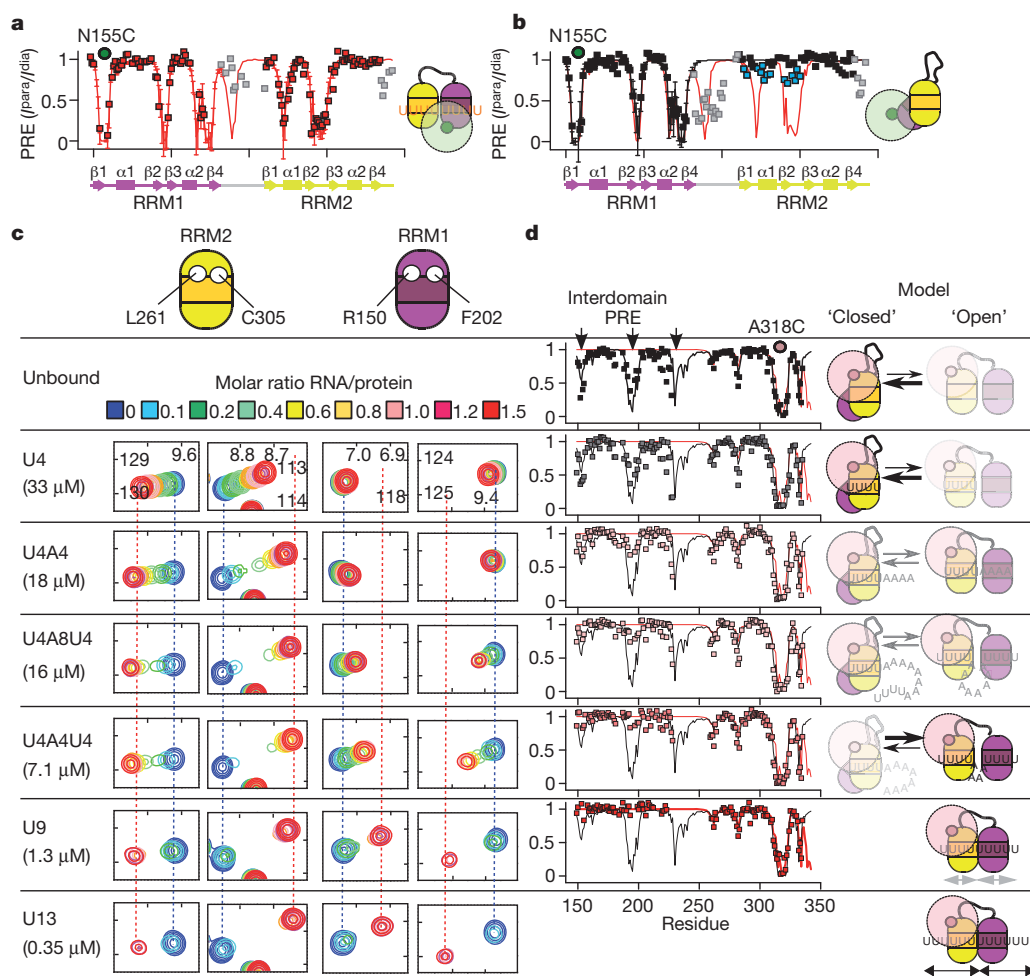
<sup>1</sup>Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany. <sup>2</sup>Institut Européen de Chimie et Biologie and Université de Bordeaux, 2 rue Robert Escarpit, 33607 Pessac, France. <sup>3</sup>Structural and Computational Biology, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. <sup>4</sup>Munich Center for Integrated Protein Science and Chair Biomolecular NMR, Department Chemie, Technische Universität München, Lichtenbergstrasse 4, 85747 Garching, Germany. <sup>5</sup>Centre de Regulació Genòmica, Universitat Pompeu Fabra, Dr. Aiguader 88, 08003 Barcelona, Spain. <sup>6</sup>Institució Catalana de Recerca i Estudis Avançats, Dr. Aiguader 88, 08003 Barcelona, Spain.

line-broadening of the NMR signals depending on the distance from the spin label (see for example Fig. 2a, b, d for spin labels at residues 155 and 318, respectively). Notably, the PRE data for the RNA-free sample indicate the presence of a pre-existing, minor population of RRM1–RRM2 that corresponds to the open form (blue squares in Fig. 2b and Supplementary Fig. 9b). This indicates that conformational sub-states resembling the open conformation of RRM1–RRM2 exist already in the absence of RNA ligand. Although similar observations have been described for other systems<sup>14–19</sup>, an equilibrium between two distinct open and closed states for binding of multi-domain proteins to a degenerate ligand motif has not been reported. We thus wondered whether Py tract recognition by U2AF65 may involve a graded shift in a pre-existing ‘multi-domain’ equilibrium by conformational selection and thus provide the molecular rationale linking the wide variety of intron RNA Py tract sequences to their encoded ‘strength’ of splicing efficiency.

To this end, we examined whether a dynamic equilibrium between the open and closed forms of U2AF65 RRM1–RRM2 could provide a mechanism for regulating the extent of Py tract binding. In the absence of RNA (closed conformation) only the RNA-binding surface of RRM2 is freely accessible for initial interactions with RNA. Thus, short RNA ligands that can cover only a single RRM domain (such as a four-uridine RNA) should bind preferentially to RRM2 and fail to alter the domain rearrangement. Consistent with this, titration of RRM1–RRM2 with U4 RNA shows significant chemical shift perturbations

only for residues in RRM2 (Fig. 2c) and the binding affinity of U4 to RRM1–RRM2 is comparable to the interaction with the isolated RRM2 (Supplementary Table 3). Moreover, the pattern of inter-domain PRE data and therefore the relative domain arrangement is very similar to that of the unbound RRM1–RRM2 (Fig. 2d and Supplementary Fig. 10). This indicates that U4 mainly binds to RRM2, and that RRM1–RRM2 is predominantly in the ‘closed’ conformation when bound to U4.

We then investigated the conformation of RRM1–RRM2 upon binding to a series of RNA ligands, representing Py tracts of various length and composition, mimicking the degeneracy of Py tracts found in human U2 introns<sup>1,5–8</sup>. Using isothermal titration calorimetry (ITC), NMR chemical shift perturbation and PRE measurements, we found that the ligands U4A4, U4A8U4 and U4A4U4 show intermediate but gradually increasing affinity to RRM1–RRM2, in comparison to the low-affinity U4 and the high-affinity U9/U13 ligands (Fig. 2c, d). Surprisingly, each Py tract RNA shows similar binding to RRM2 regardless of the overall binding affinity (Fig. 2c; comparable chemical shift perturbation for RRM2 residues). Instead, the overall increase in affinity reflects an increasing contribution of RRM1 bound to RNA, as shown by the extent of chemical shift perturbation seen for residues in RRM1 (Fig. 2c). Full binding of RRM1 appears only with a long uninterrupted stretch of pyrimidine, such as with U9 or U13, with adjacent high-affinity binding sites for both RRM1 and RRM2, and an overall affinity approaching the product of the U4 RNA affinities



**Figure 2 | Binding of Py tracts of different strength to U2AF65 RRM1–RRM2.** **a, b**, Paramagnetic relaxation enhancement (PRE) data from a spin label attached to residue 155 (green circle) for **(a)** U9-bound RRM1–RRM2 (red squares) and **(b)** unbound RRM1–RRM2 (black squares), with back-calculated PRE values (red line, U9 bound; black line, unbound) and derived from the structure ensembles (s.d. from mean). Flexible regions and minor open population data are shown by grey and blue squares, respectively. **c**, RRM1–

RRM2 titration NMR spectra with model Py tracts (dissociation constants from Supplementary Table 3 in parentheses);  $^1\text{H}/^{15}\text{N}$  chemical shifts in parts per million are shown on the x/y-axis. **d**, Experimental PRE data (peak intensities in the paramagnetic and diamagnetic state ( $I^{\text{para}}/I^{\text{dia}}$ ) versus residue number) and back-calculations for unbound and the various RRM1–RRM2–RNA ligand complexes for a spin label attached to residue 318 (pink circle) and a corresponding schematic of the equilibrium between open and closed states.

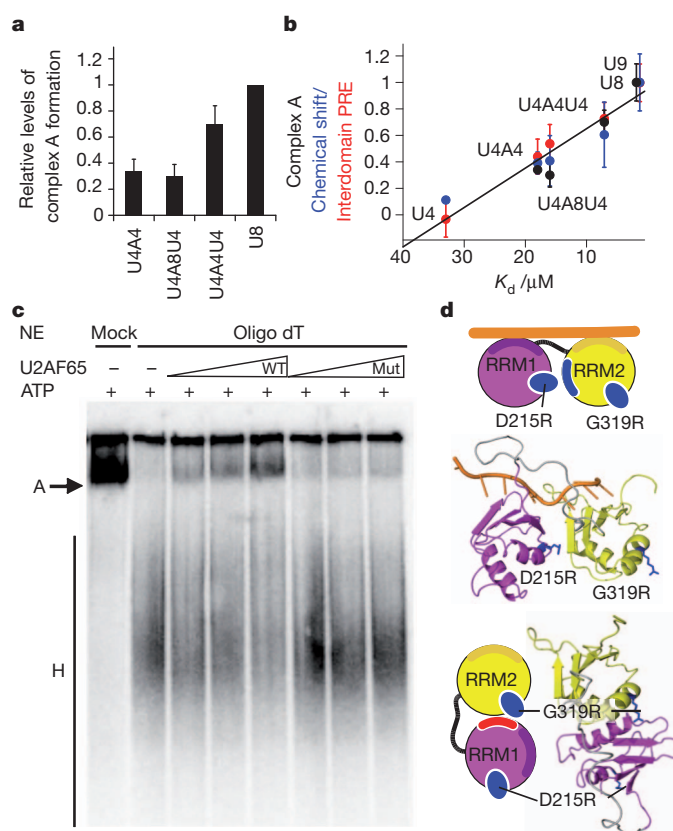


of the individual domains. Analysis of the PRE data shows a gradual change in the pattern of inter-domain PRE from the unbound form (and U4-bound) to the fully bound form (Fig. 2d and Supplementary Fig. 10a). These complexes are still mainly composed of compact states, where the two domains interact and reorient together in solution, as shown by NMR relaxation data (Supplementary Fig. 10b). As the NMR data report on population-weighted averages of the molecules in solution, it is reasonable to assume that RRM1–RRM2 exists in equilibrium between the two conformations, corresponding to the open (bound) and closed (unbound) RRM1–RRM2 structures, respectively. Therefore, the data shown in Fig. 2 indicate that binding of Py tracts of increasing affinity (or strength) results in a shift of populations between the closed and open conformations (Supplementary Fig. 10c–e). More importantly, the data indicate that RRM1 is a key regulator of this mechanism governed by the competition between binding RRM2 and binding a secondary RNA site within the Py tract.

To analyse whether this population shift from the closed to open conformation is coupled to the functional U2AF activity during spliceosome assembly, we measured U2 snRNP recruitment (pre-spliceosome (A) complex formation) on RNAs containing the 3'-splice-site region and downstream exon of adenovirus major late (AdML) promoter transcripts with both native (U8) and the different Py tract configurations analysed above (U4A4, U4A8U4, U4A4U4). Complex A is formed most efficiently with the U8 Py tract, somewhat less with U4A4U4, and with significantly lower efficiency using the U4A8U4 and U4A4 substrates (Fig. 3a and Supplementary Fig. 11a, b). There is a notable quantitative correlation between the extent of U2 snRNP recruitment, RNA binding affinity and the population of molecules adopting the open conformation of U2AF65 (Fig. 3b and Supplementary Fig. 11c). The similarity between the U4A8U4 and U4A4 substrates suggests that an eight-adenosine spacing between two consecutive uridine stretches is unable to compete for RRM1 binding against the RRM1–RRM2 interaction present in the closed conformation. Notably, the results obtained using the model RNA ligands also extend to native Py tracts represented by four human intron sequences that contain comparable length and branch-point strength (Supplementary Fig. 12).

We next designed mutants of RRM1–RRM2 with the aim to shift the equilibrium between open and closed states, thereby perturbing the degree of conformational sampling and thus affecting the formation of complex A accordingly. Several mutations were created remote from the RNA-binding surface (Supplementary Fig. 13) and investigated using ITC (Supplementary Table 3). The double mutation D215R/G319R destabilizes the open conformation (D215R) by electrostatic repulsion, whereas it strengthens the interface of the closed conformation with favourable charge complementarity (G319R) (Fig. 3d). This mutant shows reduced affinity for U9 and U4A8U4 consistent with a shift of the conformational equilibrium towards the closed state (Supplementary Fig. 14) and shows strongly reduced formation of complex A (Fig. 3c). A second mutant, RRM1–RRM2( $\Delta$ 233–252), was designed to selectively prevent the closed conformation due to a strategic shortening of the linker connecting RRM1 and RRM2. As predicted, RRM1–RRM2( $\Delta$ 233–252) favours the open conformation even in the absence of RNA ligand, as confirmed by the pattern of PRE (Supplementary Fig. 15). In addition, this mutant has increased binding to U9, U4A4U4 and U4A8U4 ligands. It displays a consistent level of binding by RRM1 in keeping with removal of the competition between binding RRM2 and RNA and shows activity in splicing assays comparable to the wild-type protein (Supplementary Fig. 15). The predicted opposing effects of these two mutants provide further support for the functional significance of the conformational equilibrium of the U2AF65 tandem RRM domains.

Our results indicate that the tandem RRM domains of U2AF65 do not simply act as a binding scaffold but instead have an active role in quantitatively relating Py tract strength to splice site recognition and spliceosome assembly (Fig. 4 and Supplementary Fig. 16). Multi-domain



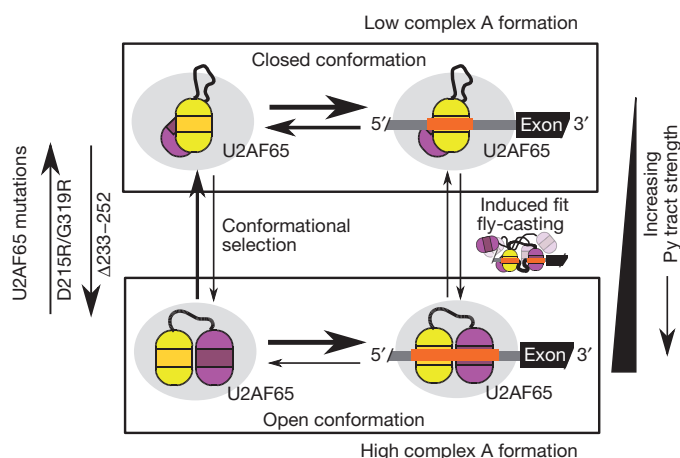
**Figure 3 | Spliceosome assembly as a function of Py tract strength.**

**a**, Complex A formation for AdML promoter transcripts with various Py tract sequences (Supplementary Fig. 10). Error bars indicate mean  $\pm$  s.d. for 11 replicates. **b**, Correlation of binding affinity with: complex A formation (black; from **a**); relative inter-domain PRE effect from the open conformation population (red; spin label at 318, Fig. 2d; error bars from 100 iterations of a Monte Carlo analysis) and relative average chemical shift perturbation of RRM1 (blue), with error bars for mean  $\pm$  s.d. The black line represents a linear fit of the data. **c**, Complex A formation in U2AF-depleted nuclear extracts with recombinant purified GST–U2AF65 (WT) or mutant D215R/G319R (Mut). Spliceosomal complexes A and H are indicated by 'A' and 'H' on the left. **d**, Design rationale for the D215R/G319R mutant.

conformational selection of the open states allows the tandem RRM domains to function as a molecular rheostat with regard to U2AF activity during early steps of splicing, involving a competition for RRM1 between binding RRM2 (autoinhibition in a closed conformation) and RNA (activation by an open conformation). This provides a selectivity filter against promiscuous RNA binding and spliceosome assembly, as the higher affinity Py tract ligands are better able to counteract the energetic penalty needed for both RRM domains to bind.

Our data do not rule out the existence of a minor induced fit mechanism involving 'fly-casting'<sup>20</sup> where the tandem RRM domains may be able to identify weak Py tracts where short pyrimidine stretches are distributed over a longer RNA sequence. After initial binding of a short (that is, four-nucleotide) pyrimidine stretch to RRM2, neighbouring uridine stretches can be screened by RRM1 to find a complete 8-mer Py tract with increased U2AF affinity. The search space of RRM1 is restricted by the conserved length (not the sequence) of the linker connecting RRM1 and RRM2 (Supplementary Fig. 8). In addition, depending on the separation of the two U4 stretches, the entropy loss associated with binding of the RNA to RRM1–RRM2 decreases as the RNA linker is shortened. This will affect the relative contribution of induced fit compared to conformational selection (Supplementary Fig. 16)<sup>21</sup>.

The tunable conformational shift described here can contribute to overall 3'-splice-site recognition beyond simply improving U2AF RNA occupancy. The open conformation may expose protein functionalities that are occluded in the closed RRM1–RRM2 state, and may thereby



**Figure 4 | Py tract recognition by U2AF.** A multi-domain conformational-selection mechanism enables Py tracts of increasing strength to capture the open conformation of U2AF and support efficient assembly of complex A. Protein mutations can shift the equilibrium to favour either the open or closed conformation (left). Relative sizes of the on- and off-rates are indicated by the thickness of the arrows (see Supplementary Text and Supplementary Fig. 16 for further details). Fly-casting may represent a minor mechanism of induced fit based on the extent of spatial separation of Py tract elements.

facilitate U2 snRNP recruitment. Most notable is the conserved  $\alpha$ -helical surface of RRM2 that is only accessible in the open orientation (Fig. 1a, b). This region contains a lysine residue (K276) that upon hydroxylation alters the splicing pattern of some genes<sup>22</sup>. The equilibrium between open and closed conformations might therefore orchestrate a distinct ribonucleoprotein assembly characteristic of activated 3' splice sites. Reciprocally, additional protein-binding partners of U2AF65 (for example, U2AF35 or others<sup>23</sup>), through additional interactions with 3'-splice-site components (for example, the AG dinucleotide), could favour the open conformation and thereby enhance the recognition of weak Py tracts.

We expect that similar mechanisms of multi-domain conformational selection coupled to biological activity operate in many multi-domain proteins that must functionally distinguish degenerate nucleotide or amino acid motifs from similar, nonspecific sequences. As demonstrated here for U2AF65, structural analysis of multiple domains connected by flexible linkers critically depends on the use of solution techniques in a multidisciplinary approach.

## METHODS SUMMARY

Wild-type and mutated U2AF65 constructs were cloned, expressed in *Escherichia coli* and purified as described in Methods. Oligoribonucleotides were purchased from Biospring GmbH. NMR spectra were collected at 295 K, with chemical shifts assigned by standard experiments or by comparison to previous data<sup>11</sup>. Residual dipolar coupling used partial alignment by Pf1 phage or a liquid crystal containing hexanol and pentaethylene glycol monododecyl ether<sup>24</sup>. NMR spectroscopy and structure calculation details are provided in Methods and Supplementary Information. Protein–RNA affinity was measured by isothermal titration calorimetry. *In vitro* assay of complex A assembly was carried out as described previously<sup>25</sup>; the splicing activity of U2AF65 mutants used recombinant protein and nuclear extracts, in which U2AF was depleted by oligo-dT cellulose chromatography<sup>26</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 9 July 2010; accepted 5 May 2011.

Published online 13 July 2011.

1. Zamore, P. D., Patton, J. G. & Green, M. R. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609–614 (1992).
2. Banerjee, H., Rahn, A., Davis, W. & Singh, R. Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF65) recognize polypyrimidine tracts using multiple modes of binding. *RNA* **9**, 88–99 (2003).
3. Banerjee, H. *et al.* The conserved RNA recognition motif 3 of U2 snRNA auxiliary factor (U2AF65) is essential *in vivo* but dispensable for activity *in vitro*. *RNA* **10**, 240–253 (2004).

4. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
5. Reed, R. The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.* **3**, 2113–2123 (1989).
6. Roscigno, R. F., Weiner, M. & Garcia-Blanco, M. A. A mutational analysis of the polypyrimidine tract of introns. Effects of sequence differences in pyrimidine tracts on splicing. *J. Biol. Chem.* **268**, 11222–11229 (1993).
7. Singh, R., Valcárcel, J. & Green, M. R. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**, 1173–1176 (1995).
8. Coolidge, C. J., Seely, R. J. & Patton, J. G. Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res.* **25**, 888–896 (1997).
9. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Struct. Mol. Biol.* **14**, 1025–1040 (2007).
10. Seet, B. T., Dikic, I., Zhou, M. M. & Pawson, T. Reading protein modifications with interaction domains. *Nature Rev. Mol. Cell Biol.* **7**, 473–483 (2006).
11. Ito, T., Muto, Y., Green, M. R. & Yokoyama, S. Solution structures of the first and second RNA-binding domains of human U2 small nuclear ribonucleoprotein particle auxiliary factor (U2AF(65)). *EMBO J.* **18**, 4523–4534 (1999).
12. Sickmier, E. A. *et al.* Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol. Cell* **23**, 49–59 (2006).
13. Simon, B., Madl, T., Mackereth, C. D., Nilges, M. & Sattler, M. An efficient protocol for NMR-spectroscopy-based structure determination of protein complexes in solution. *Angew. Chem. Int. Edn Engl.* **49**, 1967–1970 (2010).
14. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chem. Biol.* **5**, 789–796 (2009).
15. Zhang, Q., Stelzer, A. C., Fisher, C. K. & Al-Hashimi, H. M. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* **450**, 1263–1267 (2007).
16. Lange, O. F. *et al.* Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **320**, 1471–1475 (2008).
17. Li, P., Martins, I. R., Amarasinghe, G. K. & Rosen, M. K. Internal dynamics control activation and activity of the autoinhibited Vav DH domain. *Nature Struct. Mol. Biol.* **15**, 613–618 (2008).
18. Korzhnev, D. M., Religa, T. L., Banachewicz, W., Fercht, A. R. & Kay, L. E. A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329**, 1312–1316 (2010).
19. Henzler-Wildman, K. A. *et al.* A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913–916 (2007).
20. Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl Acad. Sci. USA* **97**, 8868–8873 (2000).
21. Hammes, G. G., Chang, Y.-C. & Oas, T. G. Conformational selection or induced fit: a flux description of reaction mechanism. *Proc. Natl Acad. Sci. USA* **106**, 13737–13741 (2009).
22. Webby, C. J. *et al.* Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science* **325**, 90–93 (2009).
23. Soares, L. M., Zanier, K., Mackereth, C., Sattler, M. & Valcárcel, J. Intron removal requires proofreading of U2AF/3' splice site recognition by DEK. *Science* **312**, 1961–1965 (2006).
24. Rückert, M. & Otting, G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* **122**, 7793–7797 (2000).
25. Guth, S., Tange, T., Ø., Kellenberger, E. & Valcárcel, J. Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. *Mol. Cell Biol.* **21**, 7673–7681 (2001).
26. Valcárcel, J., Martínez, C. & Green, M. R. Functional analysis of splicing factors and regulators. In *mRNA Formation and Function* 31–53 (Elsevier, 1997).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank F. Gabel, M. Nilges, C. Griesinger, J. Müller and K. Scheffzek for discussions, and H. Tilgner for analysis of natural Py tract sequences. C.D.M. acknowledges support by EMBO Long Term Fellowship, ICSN and Aquitaine regional government. T.M. thanks the Austrian Science Fund (FWF) and EMBO for postdoctoral fellowships. We thank the EU NMR LSF in Frankfurt and the Bavarian NMR Centre (BNMRZ) in Munich for NMR measurement time. This work was supported by the European Commission, grants 3D Repertoire, FSG-V-RNA and NIM3 No. 226507 (M.S.), EURASNET, AICR and Fundación Marcelino Botín (J.V.).

**Author Contributions** C.D.M., S.B., K.Z. and A.G. cloned and purified native and nitroxyl-labelled proteins. C.D.M., K.Z., B.S. and T.M. collected, processed and analysed NMR spectroscopy data. C.D.M., B.S. and T.M. calculated and analysed structural ensembles. S.B. performed *in vitro* splicing assays. V.R. performed ITC. J.V. and M.S. contributed to study design. C.D.M. and M.S. wrote the paper. All authors discussed the results and commented on the manuscript.

**Author Information** The coordinates of the open RNA-bound conformation of RRM1–RRM2 and the closed conformation in the absence of RNA are deposited in the Protein Data Bank with accession codes 2YH1 and 2YH0, respectively. All structural ensembles with explicit spin labels are available from the authors upon request. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to M.S. ([sattler@helmholtz-muenchen.de](mailto:sattler@helmholtz-muenchen.de)).



## METHODS

**Cloning.** Full-length human U2AF65, as well as the truncation mutants U2AF65(RRM1–RRM2), U2AF65(RRM1) and U2AF65(RRM2), were cloned by using PCR amplification. Primers were designed to introduce NcoI and Acc65I restriction enzyme sites, to allow for directional insertion into a modified pET9d vector containing an amino-terminal His<sub>6</sub> tag followed by a tobacco etch virus (TEV) protease cleavage site. Full-length U2AF65 constructs were cloned into a modified pET9d vector containing an N-terminal GST tag. The linker deletion and site-specific mutants were created by PCR amplification with overlapping oligonucleotides containing the mutated sequence. All plasmids were verified by sequencing.

**Expression and purification.** U2AF65-derived peptides were produced in BL21(DE3) or BL21(DE3)pLysS cells using standard media or minimal M9T media supplemented with 2 g l<sup>-1</sup> [<sup>13</sup>C]glucose and/or 1 g l<sup>-1</sup> [<sup>15</sup>N]ammonium chloride. Following normal growth, cells were induced at an OD<sub>600 nm</sub> of 0.6 with 0.25 μM IPTG followed by protein expression for 16 h at 25 °C. Cells were collected by centrifugation, lysed by sonication in the presence of lysozyme and EDTA-free Complete protease inhibitor (Roche Applied Science) then resuspended in binding buffer consisting of 50 mM Tris (pH 7.5), 500 mM NaCl, 5% (v/v) glycerol and 5 mM imidazole. The sample was added to Ni<sup>2+</sup> affinity chromatography resin and washed with 20 column volumes of binding buffer followed by five column volumes of the same buffer but with 30 mM imidazole. Elution with 50 mM Tris (pH 7.5), 500 mM NaCl, 5% (v/v) glycerol and 250 mM imidazole was followed by a buffer exchange to phosphate buffered saline using a PD10 column (GE Healthcare). Removal of the His<sub>6</sub> tag with 20 μg l<sup>-1</sup> TEV protease required from 16 h to 5 days at room temperature depending on the construct. TEV protease, His<sub>6</sub> tag and uncleaved protein were removed via a second passage of the sample through Ni<sup>2+</sup> affinity chromatography resin. The eluate was concentrated to 2.5 ml using either an Amicon Ultra-15 (Millipore) or Vivaspinn 20 (Sartorius) centrifugal filter unit. Following a final buffer exchange to 20 mM sodium phosphate (pH 6.5), 50 mM NaCl, 0.1% sodium azide and 1 mM EDTA with a PD10 column the samples were concentrated to at least 0.2 mM protein. GST-tagged protein was purified using glutathione-agarose chromatography. The sample was bound to the column in 50 mM Tris (pH 8.0), 150 mM NaCl, 2 mM dithiothreitol and 1 mM EDTA, with elution using the same buffer containing 10 mM freshly reduced glutathione. RNA oligonucleotides were purchased from Biospring GmbH.

**Spin labelling.** Residues on the surface of each RRM and distant from the RNA-binding area (namely N155, A164, A171, L187, A188, T209, D273, S281, A287 and A318) were mutated individually to cysteine. The corresponding single cysteine mutant proteins were expressed and purified as described above. Before addition of 3 molar equivalents of 3-(2-iodoacetamido)-2,2,5,5-tetramethyl-1-pyrrolidinyloxy radical (iodoacetamido-PROXYL; Sigma-Aldrich) dissolved in methanol, the protein samples were completely reduced by the addition of 2 mM dithiothreitol, and extensively dialysed in 50 mM Tris (pH 8.0) and 200 mM NaCl. Following an overnight reaction in the dark at 4 °C, the modified protein was passed three times through a PD10 desalting column (GE Healthcare Life Sciences) to remove all unreacted spin label and change the buffer to 20 mM sodium phosphate (pH 6.5), 50 mM NaCl, 0.1% sodium azide and 1 mM EDTA.

**NMR spectroscopy.** All samples contained 0.2 to 0.8 mM protein in 20 mM sodium phosphate (pH 6.5), 50 mM NaCl with 10% <sup>2</sup>H<sub>2</sub>O added for the lock. Spectra were recorded at 295 K using DRX500, DRX600, AV800 or AV900 Bruker NMR spectrometers, equipped with cryogenic triple resonance gradient probes. Spectra were processed using NMRPipe/Draw<sup>27</sup> and analysed using Sparky 3 (T. D. Goddard and D. G. Kneller, University of California) and NMRView<sup>28</sup>. Protein backbone assignments were obtained from HNCACB and HNCA spectra, or by comparison to related <sup>1</sup>H,<sup>15</sup>N-HSQC and -TROSY spectra and previously published data<sup>11</sup>. Amino acid side chain resonance assignments were obtained from standard HCCH-TOCSY, <sup>15</sup>N- and <sup>13</sup>C-edited NOESY-HSQC experiments. About 15 intermolecular NOEs between the U9 RNA and U2AF65(RRM1–RRM2) were identified for well-resolved peaks in the 3D <sup>13</sup>C-edited NOESY-HSQC experiments. For 13 of these peaks, chemical shifts of the corresponding protein signals could be assigned.

Amide <sup>15</sup>N relaxation data were acquired at 600 MHz and 295 K as described<sup>29</sup>. Steady-state heteronuclear {<sup>1</sup>H}<sup>15</sup>N-NOE spectra were recorded with and without 3 s of <sup>1</sup>H saturation. Relaxation rates and error calculations were determined using NMRView v.4 (ref. 28).

<sup>1</sup>H,<sup>15</sup>N residual dipolar couplings were measured using an interleaved spin-state-selective <sup>1</sup>H,<sup>15</sup>N-TROSY experiment. <sup>15</sup>N-<sup>13</sup>C' residual dipolar couplings were measured using a 3D-HNCO experiment<sup>30</sup>. Alignment media consisted of Pfl phage (Profos AG) or a liquid crystalline mixture of hexanol and pentaethylene glycol monododecyl ether<sup>24</sup>.

Paramagnetic relaxation enhancements (PREs) arising from the spin label were determined using a ratio of peak intensities in the paramagnetic and diamagnetic state ( $I^{para}/I^{dia}$ ) from <sup>1</sup>H,<sup>15</sup>N-HSQC and/or -TROSY spectra without and with the addition of 6 molar equivalents of ascorbic acid. RNA-bound samples included the addition of 1.5 molar equivalents of 10 mM RNA dissolved in water (BioSpring GmbH)<sup>13</sup>. In the case of the N155C mutant, the PRE was also determined directly through the measurement of <sup>1</sup>H<sup>N</sup> T<sub>1</sub> and T<sub>2</sub> relaxation times<sup>13,31</sup>. A spin label was also incorporated onto an oligoribonucleotide consisting of a 5' 4-thiouridine followed by eight standard uridine residues (BioSpring GmbH), using the same reagent and strategy as detailed above<sup>32</sup>.

**Structure calculation.** Structures were calculated using modified CNS protocols in the ARIA/CNS setup<sup>13,33,34</sup>. In brief, the protocol consists of the following steps: (1) local refinement of the available domain structures of RRM1 and RRM2 using RDC data measured from two alignment media; (2) generation of linker and spin labels, randomization of the linker residues in the RRM1-linker-RRM2 sequence; (3) molecular dynamics simulated annealing restraining RRM1 and RRM2 harmonically to their refined starting structures, with additional dihedral angle restraints from secondary chemical shifts using TALOS<sup>35</sup>, RDCs (omitted for free U2AF65) and hydrogen bond restraints.

Major changes to the standard structure calculation set-up include the generation of the template structures and the randomization protocol: the template structure is generated by reading in available domain crystal structures of RRM1 and RRM2 (ref. 12), which are then fixed by a harmonic energy potential during the simulated annealing protocol. Randomization is restricted to linker residues connecting the two RRM domains, which are kept rigid. The spin label groups are attached to cysteine residues using a patch, which allows the incorporation of one or several (non-interacting) copies of the proxyl moiety to each site. Calculations are performed with an ensemble of four spin labels per cysteine. Simulated annealing protocols and temperature course are the same as in standard structure calculations. The resulting structure ensembles were further refined by replacing the spin-labelled cysteines with the corresponding wild-type residues followed by energy minimization and final refinement in a shell of water molecules<sup>13</sup>.

For residual dipolar couplings, the structures of both RRM domains are refined individually with an effective energy constant for the positional restraints (10 kcal mol<sup>-1</sup> Å<sup>-2</sup>) allowing for local refinement of the protein backbone by the RDC restraints. This step allows slight rearrangements for the backbone atoms of some residues and improves the overall agreement with the RDC data. During the structure calculation of the tandem domain complex, these locally refined structures are then restrained with a very high effective energy constant (non-crystallographic force constant 10,000 kcal mol<sup>-1</sup> Å<sup>-2</sup>). Note that the relative domain orientation of RRM1 and RRM2 does not change if locally refined or unrefined structures are used in the protocol.

Measured intensity ratio from HSQC spectra of oxidized and reduced spin-labelled proteins were converted into paramagnetic relaxation rates and distances as described<sup>13</sup>.

Quality factors for RDC and PRE restraints are calculated as

$$Q = \sqrt{\frac{\sum (V_{\text{backcalc}} - V_{\text{exp}})^2}{\sum (V_{\text{exp}})^2}}$$

where  $V_{\text{backcalc}}$  and  $V_{\text{exp}}$  are the back calculated and experimental RDC of PRE values for a given structure.

The structural statistics for the structure ensembles with the spin-label molecules still attached are provided in Supplementary Table 2. The structural statistics for the final water-refined closed and the open, U8-bound conformations are given in Supplementary Table 1. The final ensemble of ten structures of the RNA-bound open conformation has 94.6% and 5.2% in the most favoured and additional allowed regions, respectively, for residues within RRM1 (150–229) and RRM2 (260–336). For the ensemble of ten structures of the closed conformation, the same ranges display 93.3% and 6.7% of residues in the most favoured and additional allowed regions, respectively.

**RNA titrations by NMR.** For each RNA titration, samples initially contained 0.2 mM protein in 500 μl of 20 mM sodium phosphate (pH 6.5), 50 mM NaCl, 1% sodium azide and 1 μM EDTA. Chemical shift perturbation was followed by measuring <sup>1</sup>H,<sup>15</sup>N-HSQC and/or -TROSY-HSQC spectra with cumulative addition of 10 mM RNA (BioSpring GmbH) dissolved in H<sub>2</sub>O. Typical titration series used steps of 0, 0.1, 0.2, 0.4, 0.6, 0.8, 1, 1.2 and 1.5 molar equivalents of RNA to protein.

**Isothermal titration calorimetry.** ITC was carried out using VP-ITC or ITC200 Microcal calorimeters (Microcal) at 25 °C. All proteins were dialysed extensively using Slide-A-Lyzer 3.5-kDa molecular weight cutoff cassettes (Pierce Biotechnology) against 20 mM sodium phosphate (pH 6.5), 50 mM NaCl and 0.1 mM EDTA. Buffer from the dialysis was used to resolubilize the RNA (BioSpring GmbH)



and to provide a baseline as required. The data were analysed using program Origin version 5.0 provided by Microcal.

**Surface plasmon resonance.** An IAsys resonant mirror biosensor (AffinitySensors) was used to determine the equilibrium constants for the interactions of biotinylated U9 oligonucleotide with RRM1, RRM2 and RRM1–RRM2 proteins<sup>36</sup>. The cuvette was prepared by an initial capture of neutravidin on the biotin-coated surface, and subsequent attachment of the biotinylated U9 RNA. Nonspecifically bound U9 was removed by washing with 2 M NaCl followed by washes with PBS, containing 0.1% Tween20 and binding buffer (20 mM sodium phosphate at pH 6.5, 50 mM NaCl and 1  $\mu$ M EDTA). Following cuvette equilibration with binding buffer, association phase binding responses were recorded separately for various protein concentrations with subsequent washing of the cuvette and monitoring of the dissociation phase. The sensor surface was regenerated by sequential washing with 2 M NaCl and binding buffer. As a negative control for binding experiments immobilized neutravidin was used. All experiments were performed at 20 °C. The experimental data were corrected for nonspecific binding and analysed by using the FASTfit software provided by the manufacturer.

**Singular value decomposition analysis.** To determine the fraction of the unbound and bound populations of RRM1–RRM2 with the different Py tract RNA, the PRE data were fitted as a linear combination of the PRE data of the free RRM1–RRM2, and those of the U9-bound RRM1–RRM2. Only residues for which the PRE data were used as restraints in the calculation of the free and U9-bound structures, and which show inter-domain bleaching, were considered. For the spin label attached to residue 318, 51 residues in RRM1 were included in the fit, with the results expressed as the fraction of U9-bound conformation. Error analysis consisted of 100 iterations of a Monte Carlo simulation with error added only to the experimental data and not the two models.

**In vitro splicing assays.** Pre-spliceosome A complex assembly was carried out as described previously<sup>25</sup>. *In vitro* transcribed RNAs corresponded to the 3' half of intron 1 (including the 3'-splice-site region) and part of exon 2 of AdML promoter transcripts. The sequence of the wild-type (U8) is gggaagcuugcugcagcguaggcgcauguaguccaggguuuccuugauguacuuuauccuguccuuuuuuuuccacagCUCGCGG UUGAGGACAAACUCUUCGCGGUCUUUCCAGUGGGGAUCC; intron and exon nucleotides are indicated in lower- and uppercase letters, respectively, and the underlined sequence was replaced by uuuuuaaa (U4A4), uuuuuuuuuuuu (U4A4U4) or uuuuuuuuuuuuuuuu (U4A8U4) to generate the different mutant substrates. 40,000 c.p.m. (20 fmols) of each <sup>32</sup>P-UTP body-radiolabelled RNA substrate for various mutants (RNA integrity and amounts verified by denaturing gel electrophoresis) were incubated with varying amounts of HeLa cell nuclear extracts (CILBIOTECH; ATP depleted by incubation 30 min at 30 °C) supplemented with 3 mM MgCl<sub>2</sub>, 24.9 mM KCl, 3.33% PVA, 13.3 mM HEPES pH 8, 0.13 mM EDTA, 13.3% glycerol, 0.03% NP-40, 0.66 mM DTT and supplemented or not with 2 mM

ATP and 22 mM creatine phosphate in a final volume of 9  $\mu$ l. The mixture was incubated for 5 min at 30 °C (Supplementary Fig. 11a) or for different time points (Supplementary Fig. 11b). 1  $\mu$ l of heparin (10  $\mu$ g  $\mu$ l<sup>-1</sup>) was added and incubated for 10 min at room temperature. 3  $\mu$ l of 50% glycerol were added and 10  $\mu$ l loaded on a composite gel (4% acrylamide, 0.05% bis-acrylamide, 0.5% agarose, 50 mM Tris, 50 mM glycine). The gel was run for 6 h at 200 V in a cold room in 50 mM Tris, 50 mM glycine buffer. The gel was dried and exposed overnight with a PhosphorImager screen. Quantification for experiments using 3  $\mu$ l HeLa cell nuclear extracts with 5 min incubation at 30 °C was carried out using Image Quant v5.2. Complex A formation was tested in nuclear extracts depleted of U2AF by chromatography in oligo-dT cellulose<sup>26</sup> complemented with 2.5, 7 and 22 ng  $\mu$ l<sup>-1</sup> of recombinant purified GST–U2AF65 (WT) or mutant D215R/G319R (Mut). Results were reproducibly obtained with different depleted extracts and recombinant proteins harbouring different tags. Under optimal conditions of depletion and complementation, the wild-type protein was more than 2.5 times more active than the mutant protein.

27. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
28. Johnson, B. A. & Blevins, R. A. NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
29. Farrow, N. A. *et al.* Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994).
30. Yang, D., Venters, R. A., Mueller, G. A., Choy, W. Y. & Kay, L. E. TROSY-based HNCO pulse sequences for the measurement of 1HN-15N, 15N-13CO, 1HN-13CO, 13CO-13C $\alpha$  and 1HN-13C $\alpha$  dipolar couplings in 15N, 13C, 2H-labelled proteins. *J. Biomol. NMR* **14**, 333–343 (1999).
31. Iwahara, J., Schwieters, C. D. & Clore, G. M. Ensemble approach for NMR structure refinement against <sup>1</sup>H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J. Am. Chem. Soc.* **126**, 5879–5896 (2004).
32. Ramos, A. & Varani, G. A new method to detect long-range protein-RNA contacts: NMR detection of electron-proton relaxation induced by nitroxide spin-labeled RNA. *J. Am. Chem. Soc.* **120**, 10992–10993 (1998).
33. Nilges, M. Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.* **245**, 645–660 (1995).
34. Brünger, A. T. *et al.* Crystallography 1 NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
35. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
36. Buckle, P. E. *et al.* The resonant mirror: a novel optical sensor for direct sensing of biomolecular interactions part II: applications. *Biosens. Bioelectron.* **8**, 355–363 (1993).

# Protein targeting and degradation are coupled for elimination of mislocalized proteins

Tara Hessa<sup>1</sup>, Ajay Sharma<sup>1</sup>, Malaiyalam Mariappan<sup>1</sup>, Heather D. Eshleman<sup>1†</sup>, Erik Gutierrez<sup>1,2</sup> & Ramanujan S. Hegde<sup>1†</sup>

A substantial proportion of the genome encodes membrane proteins that are delivered to the endoplasmic reticulum by dedicated targeting pathways<sup>1</sup>. Membrane proteins that fail targeting must be rapidly degraded to avoid aggregation and disruption of cytosolic protein homeostasis<sup>2,3</sup>. The mechanisms of mislocalized protein (MLP) degradation are unknown. Here we reconstitute MLP degradation *in vitro* to identify factors involved in this pathway. We find that nascent membrane proteins tethered to ribosomes are not substrates for ubiquitination unless they are released into the cytosol. Their inappropriate release results in capture by the Bag6 complex, a recently identified ribosome-associating chaperone<sup>4</sup>. Bag6-complex-mediated capture depends on the presence of unprocessed or non-inserted hydrophobic domains that distinguish MLPs from potential cytosolic proteins. A subset of these Bag6 complex 'clients' are transferred to TRC40 for insertion into the membrane, whereas the remainder are rapidly ubiquitinated. Depletion of the Bag6 complex selectively impairs the efficient ubiquitination of MLPs. Thus, by its presence on ribosomes that are synthesizing nascent membrane proteins, the Bag6 complex links targeting and ubiquitination pathways. We propose that such coupling allows the fast tracking of MLPs for degradation without futile engagement of the cytosolic folding machinery.

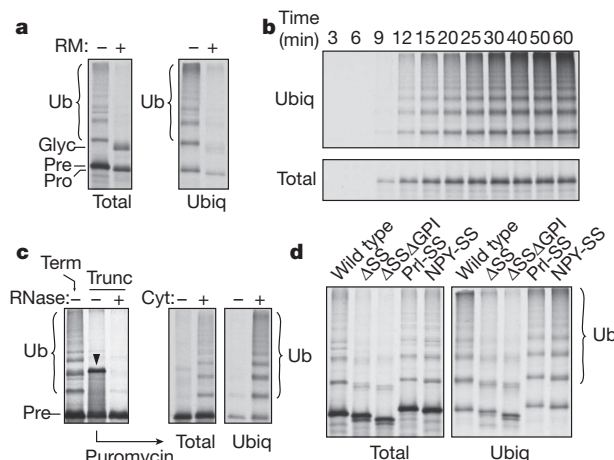
Protein targeting and translocation to the endoplasmic reticulum (ER) are not perfectly efficient<sup>5,6</sup>, thereby necessitating pathways for the degradation of MLPs that have been inappropriately released into the cytosol. For example, mammalian prion protein (PrP), a widely expressed glycosyl phosphatidylinositol (GPI)-anchored cell surface glycoprotein, displays ~5–15% translocation failure *in vitro* and *in vivo*<sup>2,3,5–10</sup>. This non-translocated population of PrP is degraded efficiently by a proteasome-dependent pathway, limiting the cytosolic PrP levels at steady state<sup>2,3,9,10</sup>. Prompt degradation is essential because mislocalized PrP can aggregate, make inappropriate interactions, and cause cell death and neurodegeneration<sup>2,11–14</sup>. The pathways for efficient disposal of MLPs, however, are not known.

To study this problem, we reconstituted the ubiquitination of mislocalized PrP *in vitro*. Radiolabelled PrP synthesized in rabbit reticulocyte lysate (RRL) supplemented with ER-derived rough microsomes was predominantly translocated into the ER, processed and glycosylated (Fig. 1a). However, various conditions that reduced the extent of translocation—such as omission of rough microsomes, inactivation of signal recognition particle (SRP)-dependent targeting or blocking of translocation through the translocon—all resulted in increased PrP ubiquitination in a lysine-dependent manner (Fig. 1a and Supplementary Figs 1–3). Other mislocalized secretory and membrane proteins were also similarly ubiquitinated in the cytosol (Supplementary Fig. 4). The ubiquitination of mislocalized PrP closely parallels PrP synthesis (Fig. 1b), suggesting that ubiquitination is rapid. Yet, ubiquitination occurred strictly post-translationally, because full-length PrP that was tethered as a nascent peptidyl-transfer RNA to the ribosome was not ubiquitinated until it had been released into the cytosol through the

action of puromycin (Fig. 1c and Supplementary Fig. 5). An unrelated membrane protein behaved similarly (Supplementary Fig. 6).

Efficient ubiquitination of PrP was strongly dependent on unprocessed hydrophobic signals at the amino and carboxy termini (Fig. 1d). Conversely, green fluorescent protein (GFP) became a substrate for ubiquitination when hydrophobic targeting signals were added (Supplementary Fig. 4). Ubiquitination was therefore not solely a consequence of protein misfolding, because PrP lacking both the N-terminal targeting signal (denoted  $\Delta$ SS) and the C-terminal GPI-anchoring signal ( $\Delta$ GPI) was misfolded owing to its lack of glycosylation and disulphide bond formation, but was poorly ubiquitinated. This finding suggested the existence of a specialized pathway for hydrophobic-domain-containing MLPs that works more rapidly than traditional quality control pathways, which engage only after repeated failures at folding<sup>15,16</sup>.

To identify factors involved in the MLP degradation pathway, we combined biochemical fractionation and functional reconstitution assays. We produced a translation-competent fractionated RRL (Fr-RRL) (Supplementary Fig. 7) that selectively decreased the ubiquitination of non-translocated PrP (Fig. 2a) and other MLPs (Supplementary



**Figure 1 | Non-translocated PrP is rapidly ubiquitinated.** **a**, The translation of radiolabelled PrP in RRL, with or without rough microsomes (RMs), was analysed directly (left) or after isolation of ubiquitinated (ubi) products (right) by using SDS-PAGE and autoradiography. Glycosylated (glyc), precursor (pre), processed (pro) and ubiquitinated (Ub) bands are indicated. **b**, Time course of PrP synthesis (bottom) and PrP ubiquitination (top) *in vitro*. **c**, PrP containing a termination codon (term) or lacking this codon (trunc) was translated *in vitro*. Truncated PrP was released using puromycin, in the absence or presence of cytosol (cyt), and total protein and ubiquitination were analysed. The arrowhead indicates tRNA-containing PrP, which can be digested by RNase. **d**, Wild-type PrP or constructs lacking the signal sequence ( $\Delta$ SS) or both the signal sequence and GPI anchor ( $\Delta$ SSAGPI) were analysed directly or after isolation of ubiquitinated products. PrL-SS and NYP-SS contain signal sequence from preprolactin and neuropeptide Y, respectively.

<sup>1</sup>Cell Biology and Metabolism Program, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>2</sup>Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218, USA. <sup>†</sup>Present addresses: Department of Physiology, University of California, San Francisco, California 94158, USA (H.D.E.); MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK (R.S.H.).

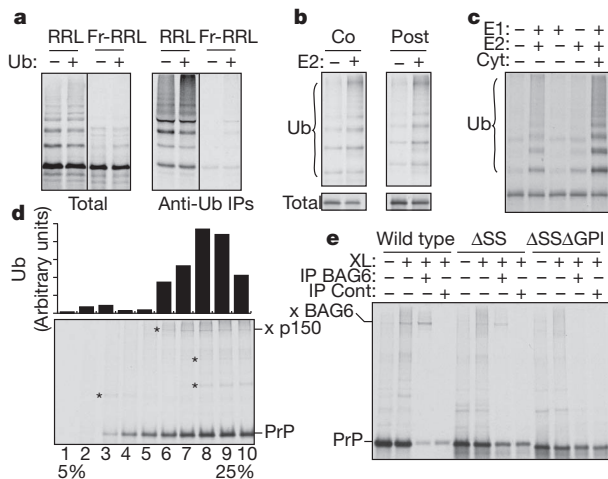
Fig. 8) but not ubiquitination in general (Supplementary Fig. 7). The missing factor in Fr-RRL (other than ubiquitin, which we included in all assays) proved to be the E2 ubiquitin-conjugating enzyme UBC5 (also known as UBE2D1) (Fig. 2b and Supplementary Figs 8 and 9). Because UBC5 restored ubiquitination equally well when added during or after PrP translation (Fig. 2b), we surmised that at least a certain population of PrP remains in a ubiquitination-competent state. Indeed, PrP and other MLPs that were affinity purified from Fr-RRL under native conditions could be ubiquitinated simply by adding purified E1, UBC5, ubiquitin and ATP (Fig. 2c and Supplementary Fig. 10).

To identify factors that maintain the ubiquitination competence of MLPs, the Fr-RRL translation products were separated by size in a sucrose gradient, and each fraction was subjected to parallel ubiquitination and chemical crosslinking analyses (Fig. 2d and Supplementary Fig. 11). The fractions retaining maximum ubiquitination competence for two different substrates correlated well with a ~150-kDa crosslinking partner (Fig. 2d and Supplementary Fig. 11). This interaction was direct (Supplementary Fig. 12) and was strongly dependent on the presence of unprocessed N- and C-terminal signals in PrP (Fig. 2e and Supplementary Fig. 13), correlating with the requirements for ubiquitination (Fig. 1d). On the basis of molecular weight, dependence on hydrophobic domains for interaction and migration position in the sucrose gradient, we surmised that the ~150-kDa crosslinked protein might be BAG6 (also called BAT3 and Scythe), a hypothesis that was subsequently verified by immunoprecipitation experiments (Fig. 2e and Supplementary Figs 13 and 14). BAG6 was recently identified as part of a three-protein ribosome-interacting chaperone complex (composed of BAG6, TRC35 and UBL4A)<sup>4</sup> that is involved in tail-anchored membrane-protein insertion into the ER<sup>4,17</sup>. A combination of crosslinking, affinity purification and immunoblotting studies verified that all three subunits of this complex are associated with MLPs

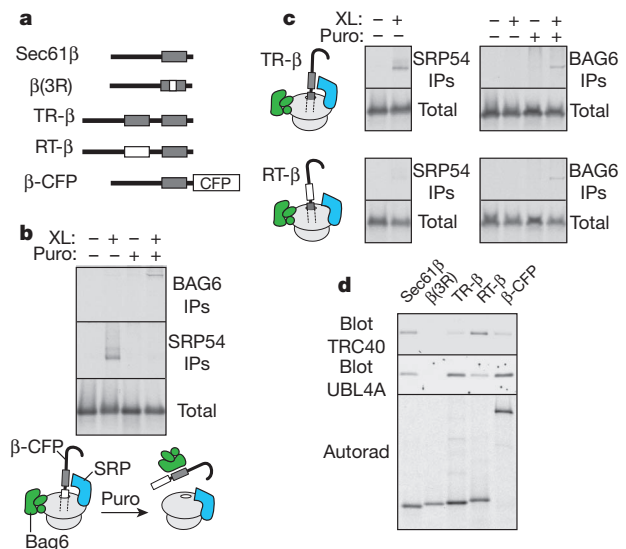
(Supplementary Figs 14 and 15, and data not shown). Thus, the Bag6 complex binds to multiple MLPs through their hydrophobic domains and has a broader specificity than only binding tail-anchored proteins.

To determine when the Bag6 complex first captures MLPs, we analysed ribosome-nascent chains (RNCs) of membrane proteins. When a transmembrane domain (TMD) emerged from the ribosomal 'tunnel', a direct interaction with SRP54 (the signal-sequence-binding subunit of the SRP) was detected by crosslinking experiments (Fig. 3a–c). By contrast, the Bag6 complex, even though it has been found to reside on such RNCs and is abundant in the cytosol<sup>4</sup>, did not make direct contact with the substrate (Fig. 3b, c). When the TMD was still inside the ribosomal tunnel, the RNC was not crosslinked to either BAG6 or SRP54 (Fig. 3c), even though both complexes can be recruited to such ribosomes<sup>4,18</sup>. After puromycin release of each of these RNCs (with the TMD inside versus outside the ribosomal tunnel), BAG6 crosslinking was observed (Fig. 3b, c). Thus, the Bag6 complex captures substrates concomitant with or after the release of nascent chains from the ribosome; these same hydrophobic domains are bound by the SRP as long as the TMD is exposed as an RNC<sup>19</sup>.

Earlier analysis of tail-anchored and non-tail-anchored membrane proteins had shown that only tail-anchored membrane proteins are efficiently loaded onto TRC40 (also known as ASNA1), the targeting factor for tail-anchored protein insertion into the ER<sup>20</sup>. Indeed, modifying a tail-anchored protein either by placing cyan fluorescent protein (CFP) polypeptide sequences after the TMD (a construct denoted  $\beta$ -CFP) (Fig. 3a) or by adding an extra TMD (denoted TR- $\beta$ ) reduced the interactions with TRC40 and simultaneously increased the interactions with the Bag6 complex (Fig. 3d). Similarly, comparison of the crosslinking partners of PrP and those of the tail-anchored protein Sec61 $\beta$  showed that both of these proteins interact with the Bag6 complex, but only Sec61 $\beta$  is primarily found bound to TRC40 (Supplementary Fig. 15). Given that the loading of tail-anchored proteins onto TRC40 depends on the Bag6 complex<sup>4</sup>, these data suggest that the Bag6 complex is acting as a triage



**Figure 2 | BAG6 interacts with MLPs through hydrophobic domains.** **a**, PrP translated in RRL or Fr-RRL, with or without 10  $\mu$ M ubiquitin (Ub), was analysed directly (left) or after anti-ubiquitin antibody immunoprecipitation (IP) (right) by using SDS-PAGE and autoradiography. **b**, PrP translated in Fr-RRL was ubiquitinated when UBC5 (E2; 250 nM) was included co-translationally (co) or post-translationally (post). Total synthesis (bottom) and ubiquitinated products (top) are shown. **c**, PrP was immunoaffinity purified under native conditions and incubated with the indicated components (cyt, cytosol; E1 enzyme, 100 nM; E2 enzyme, UBC5, 250 nM). All reactions contained His-ubiquitin and ATP. Purified ubiquitinated products are shown. **d**, PrP translated in Fr-RRL was separated into ten fractions in a 5–25% sucrose gradient. The fractions were subjected to chemical crosslinking (bottom) or ubiquitination assays (top). Asterisks indicate crosslinks. Histogram bars indicate the amount of ubiquitinated product in each fraction. The ~150-kDa crosslinking partner (x p150) is indicated. **e**, Crosslinking reactions (XL) of *in vitro*-synthesized PrP or PrP deletion constructs were analysed directly or after immunoprecipitation with anti-BAG6 or control (cont) antibodies. The crosslink to BAG6 (x BAG6) is indicated.



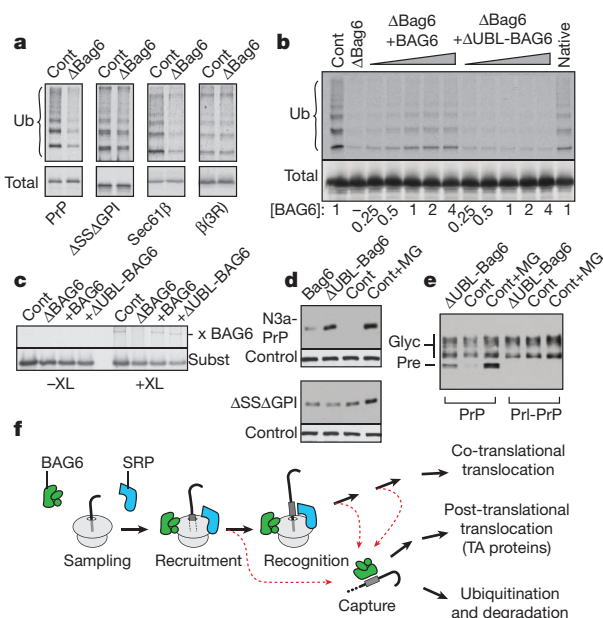
**Figure 3 | BAG6 captures MLPs released from the ribosome.** **a**, Diagram of constructs derived from Sec61 $\beta$ , with transmembrane domains shown as grey boxes and hydrophilic changes in white boxes. **b**, RNCs of  $\beta$ -CFP with the TMD outside the ribosome were subjected to crosslinking (XL) before or after release by puromycin (puro) and were analysed directly (bottom) or after immunoprecipitation (IP) with anti-BAG6 antibody (top) or anti-SRP54 antibody (centre). The results are also illustrated diagrammatically: Bag6 complex, green; SRP, blue; and ribosome, pale grey. **c**, The assays were as described in **b** but using TR- $\beta$  (top) and RT- $\beta$  (bottom). **d**, The indicated constructs were translated *in vitro*, immunoaffinity purified through their N terminus, and immunoblotted with anti-TRC40 antibody or anti-UBL4A antibody (the latter to detect the Bag6 complex). The autoradiograph shows equal recovery of the translated substrates.



factor: that is, it captures a relatively broad range of membrane proteins after their ribosomal release but transfers only a subset of them (namely, tail-anchored proteins) to TRC40 for post-translational membrane insertion. The remainder seem to be targeted for ubiquitination because of their persistent interaction with BAG6.

To examine this hypothesis, we immunodepleted the Bag6 complex from RRL (Supplementary Fig. 16) and found that the ubiquitination of several MLPs was reduced (Fig. 4a and Supplementary Fig. 17). By contrast, the control protein GFP was not ubiquitinated in RRL but became a substrate when it was attached to either a ubiquitin molecule or any of several hydrophobic ER-targeting domains (Supplementary Fig. 18). Only the hydrophobically modified GFP proteins were BAG6 dependent in their ubiquitination, consistent with their interaction with BAG6 by crosslinking analysis (Supplementary Fig. 13). Conversely,  $\Delta$ SS $\Delta$ GPI-PrP, which does not interact with BAG6 (Fig. 2e), was ubiquitinated (albeit slowly and less efficiently) in a BAG6-independent manner (Fig. 4a). Disrupting the TMD of Sec61 $\beta$  with three arginine residues (denoted  $\beta$ (3R)), which disrupts BAG6 interaction<sup>4</sup>, also resulted in less ubiquitination, which was no longer BAG6 dependent (Fig. 4a). Thus, the Bag6 complex is not required for ubiquitination of all misfolded proteins but is especially important for the efficient ubiquitination of MLPs.

When recombinant BAG6 (Supplementary Fig. 16) was added to translation extracts that had been depleted of the Bag6 complex, the ubiquitination of a model MLP was restored (Fig. 4b), and the recombinant BAG6 interacted with this MLP in crosslinking assays (Fig. 4c).



**Figure 4 | Maximum ubiquitination of MLPs requires BAG6.** **a**, Various constructs (listed at bottom) were assayed for ubiquitination in lysates containing Bag6 complex (control, cont) or lacking Bag6 complex ( $\Delta$ Bag6). The gels for assessing ubiquitination for the  $\Delta$ SS $\Delta$ GPI and  $\beta$ (3R) constructs were exposed about threefold longer than those for PrP and Sec61 $\beta$ . **b**, Bag6-complex-depleted lysates ( $\Delta$ Bag6) were replenished with increasing amounts (wedges) of recombinant BAG6 (Supplementary Fig. 16),  $\Delta$ UBL-BAG6 or native Bag6 complex and then analysed for the ubiquitination of TR- $\beta$ . Relative BAG6 levels are indicated (listed at bottom). **c**, TR- $\beta$  interacts with recombinant BAG6 and  $\Delta$ UBL-BAG6 by crosslinking (XL). Subst, substrate; x BAG6, crosslink to BAG6. **d**, The indicated PrP constructs (N3a-PrP and  $\Delta$ SS $\Delta$ GPI) were co-transfected with Bag6 complex,  $\Delta$ UBL-BAG6 complex or control plasmid (cont) (Supplementary Fig. 20), and PrP was detected by immunoblotting. One sample was treated with the proteasome inhibitor MG132 (MG) for 4 h. A loading control (control) is also shown. **e**, Effect of the  $\Delta$ UBL-BAG6 complex on wild-type PrP and Prl-PrP. Unglycosylated precursor PrP (pre) is preferentially stabilized by either overexpression of the  $\Delta$ UBL-BAG6 complex or inhibition of the proteasome. **f**, The model we propose is that the Bag6 complex captures ribosomally released hydrophobic proteins (red arrows) and triages them between post-translational targeting (for tail-anchored (TA) proteins) and ubiquitination.

BAG6 lacking its N-terminal UBL domain ( $\Delta$ UBL-BAG6) was inactive in restoring ubiquitination (Fig. 4b) despite interacting normally with substrate (Fig. 4c). This finding suggested that BAG6 may recruit the ubiquitination machinery to substrates through its UBL domain. To test this, Flag-tagged recombinant BAG6 or  $\Delta$ UBL-BAG6 was added to the Fr-RRL translation system lacking the E2 enzyme UCH5 (Supplementary Fig. 7). BAG6-substrate complexes were immunopurified through the Flag tag and incubated with purified E1 ubiquitin-activating enzyme, E2 enzyme, ubiquitin and ATP. Substrate ubiquitination was observed with BAG6 but not  $\Delta$ UBL-BAG6, verifying that the UBL domain recruits the ubiquitination machinery to the substrate (Supplementary Fig. 19). Indeed, BAG6 has been observed to interact with E3 ubiquitin ligases through its UBL domain<sup>21,28</sup>.

In Fig. 4b, c, the data indicated that  $\Delta$ UBL-BAG6 should act as a dominant negative and partly stabilize BAG6 substrates, thereby providing a selective tool for *in vivo* analysis. We therefore overexpressed the Bag6 complex or the  $\Delta$ UBL-BAG6 complex (by about twofold) (Supplementary Fig. 20) in cultured cells and assessed the levels of a co-expressed MLP substrate. A translocation-impaired signal-sequence mutant of PrP (termed N3a-PrP)<sup>5</sup> was stabilized by the  $\Delta$ UBL-BAG6 complex but almost unaffected by the wild-type Bag6 complex (Fig. 4d). Importantly,  $\Delta$ SS $\Delta$ GPI-PrP, which does not interact with BAG6 (Fig. 2e), was unaffected by either Bag6 complex or  $\Delta$ UBL-BAG6 complex overexpression (Fig. 4d) and showed higher steady-state levels than N3a-PrP (data not shown). This finding suggests that degradation is occurring by a different quality control pathway, consistent with the failure of  $\Delta$ SS $\Delta$ GPI-PrP to be recognized as an MLP (Fig. 2e).

Wild-type PrP, the translocation of which is slightly inefficient *in vivo*<sup>2,3,6,8–10</sup>, showed preferential stabilization of a non-glycosylated species when co-overexpressed with  $\Delta$ UBL-BAG6 complexes (Fig. 4e and Supplementary Fig. 21). This species was stabilized by proteasome inhibition and had been shown in earlier studies to be a non-translocated PrP precursor<sup>2,3,9,10</sup>. Replacing the slightly inefficient PrP signal sequence with the efficient signal from preprolactin (Prl-PrP) precluded the generation of non-glycosylated PrP with either proteasome inhibition or  $\Delta$ UBL-BAG6 complex overexpression (Fig. 4e). Although the extent of stabilization seems modest, it is comparable to that seen after 2 h proteasome inhibition (Supplementary Fig. 21). Partial knockdown of BAG6 with a short hairpin RNA (shRNA) similarly stabilized a non-glycosylated species of PrP (Supplementary Fig. 22). Thus, MLPs are not only generated *in vivo*<sup>2,3,6,8–10</sup>, but also require functional BAG6 for maximally efficient degradation.

Our results reveal a pathway for MLP degradation and identify an unexpectedly close link with protein targeting (Fig. 4f). Ribosomes synthesizing nascent membrane proteins can recruit both the SRP and Bag6 complex on entry of the first hydrophobic segment into the ribosomal tunnel<sup>4,18</sup>. This is a potential targeting complex for the ER membrane in both the co-translational and post-translational membrane-protein insertion pathways. We now find that such ribosomes are also potential degradation complexes because the first component of this degradation pathway is already poised to act in the event of failed targeting or inappropriate release from the ribosome. BAG6 therefore imposes a degradative fate on membrane proteins that can be avoided only by productive targeting.

Because membrane proteins would never fold in the cytosol, their direct degradation by a specialized pathway may be important to avoid unnecessarily occupying essential cellular folding pathways, particularly under conditions of stress. MLPs are distinguished from nascent cytosolic proteins by relatively long linear hydrophobic stretches, a feature that is important for BAG6 recognition. Indeed, mutagenesis shows that even modest reductions of TMD hydrophobicity sharply curtail BAG6 interaction<sup>4</sup>. This specificity distinguishes BAG6 from more general chaperones such as heat-shock protein 70 (HSP70), the substrate-binding pocket of which seems more suited to the shorter, moderately hydrophobic segments that typify nascent cytosolic proteins. This differential specificity probably explains how MLPs are triaged differently from other

potential substrates of cytosolic quality control<sup>15,16,22–28</sup>. These pathways could intersect or cooperate in as yet undefined ways given that BAG6 and HSP70 have been observed to co-immunoprecipitate<sup>26</sup>.

In addition to this role in degradation, the Bag6 complex also facilitates the loading of tail-anchored proteins onto TRC40 for post-translational insertion into the ER<sup>4</sup>. As expected, tail-anchored proteins were also ubiquitinated by way of BAG6 in the absence of, or saturation of, TRC40 (Supplementary Fig. 23). Thus, substrates of both the co-translational and post-translational targeting pathways are ubiquitinated in a BAG6-dependent manner when targeting fails. After ubiquitination, BAG6 might chaperone its polyubiquitinated substrates to the proteasome, a function that was recently proposed on the basis of the co-immunoprecipitation of BAG6 with polyubiquitinated proteins<sup>26</sup>. The Bag6 complex is therefore a multi-purpose triage factor for chaperoning especially hydrophobic proteins through the aqueous cytosol. This view conceptually links its roles in tail-anchored protein targeting<sup>4,17</sup>, in the MLP pathway (in this study), as a chaperone for newly dislocated proteins during ER-associated protein degradation<sup>27,28</sup> and in the delivery of terminally misfolded proteins to the proteasome<sup>26</sup>.

## METHODS SUMMARY

**Reagents and standard methods.** The plasmids and antibodies used and the assays carried out (*in vitro* translation assays, sucrose gradient separations, chemical crosslinking analyses, immunoprecipitation assays and immunodepletion assays) were as previously described<sup>2–8,14,20,29,30</sup>. Pull-down assays with Co<sup>2+</sup> immobilized on chelating sepharose were performed on samples that had been denatured in boiling 1% SDS and then diluted tenfold in 4 °C pull-down buffer: 0.5% Triton X-100, 25 mM HEPES, 100 mM NaCl and 10 mM imidazole. Culture, transfection and immunoblotting analysis of N2a cells (dominant-negative inhibition experiments) and HeLa cells (for shRNA experiments) were carried out as previously described<sup>2,3</sup>. Full-length BAG6 (or ΔUBL-BAG6, which lacks residues 15–89) tagged at the C terminus with a Flag epitope was overexpressed after transient transfection of HEK-293T cells and then purified with anti-Flag resin under high salt (400 mM potassium acetate) conditions.

**Modified translation extracts.** Fr-RRL contained native ribosomes (isolated from RRL) mixed with a diethylaminoethyl (DEAE) sepharose ion-exchange chromatography elution fraction prepared from ribosome-free RRL (Supplementary Fig. 7). Fr-RRL was adjusted to the following final conditions for translation: 72 mM potassium acetate, 2.5 mM magnesium acetate, 10 mM HEPES, pH 7.4, 2 mM dithiothreitol (DTT), 0.2 mg ml<sup>-1</sup> liver transfer RNA, 1 mM ATP, 1 mM GTP, 12 mM creatine phosphate, 40 μg ml<sup>-1</sup> creatine kinase, 40 μM each amino acid (except methionine) and 1 μCi ml<sup>-1</sup> [<sup>35</sup>S]methionine.

**Ubiquitination assays.** For full-length proteins, translations containing 10 μM His-tagged ubiquitin were carried out for 1 h at 32 °C. In Fr-RRL, post-translational ubiquitination was initiated by adding E2 enzyme to a final concentration of 250 nM and incubating for 1 h at 32 °C. For RNCs, samples were supplemented with E1 enzyme (85 nM), E2 enzyme (usually 250 nM or 500 nM), cytosol (RRL or Fr-RRL), 10 μM His-ubiquitin, an ATP-regenerating system (1 mM ATP, 10 mM creatine phosphate and 40 μg ml<sup>-1</sup> creatine kinase) and 1 mM puromycin. The reaction conditions were 100 mM potassium acetate, 50 mM HEPES, pH 7.4, 5 mM MgCl<sub>2</sub> and 1 mM DTT. Incubations were carried out for 1 h at 32 °C. On-bead ubiquitination of affinity-purified products was carried out under the same conditions, except without the inclusion of puromycin.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 8 December 2010; accepted 6 May 2011.**

**Published online 10 July 2011.**

1. Cross, B. C., Sinning, I., Lührink, J. & High, S. Delivering proteins for export from the cytosol. *Nature Rev. Mol. Cell Biol.* **10**, 255–264 (2009).
2. Rane, N. S., Yonkovich, J. L. & Hegde, R. S. Protection from cytosolic prion protein toxicity by modulation of protein translocation. *EMBO J.* **23**, 4550–4559 (2004).
3. Kang, S. W. *et al.* Substrate-specific translocation attenuation during ER stress defines a pre-emptive quality control pathway. *Cell* **127**, 999–1013 (2006).
4. Mariappan, M. *et al.* A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature* **466**, 1120–1124 (2010).
5. Kim, S. J., Mitra, D., Salerno, J. R. & Hegde, R. S. Signal sequences control gating of the protein translocation channel in a substrate-specific manner. *Dev. Cell* **2**, 207–217 (2002).

6. Levine, C. G., Mitra, D., Sharma, A., Smith, C. L. & Hegde, R. S. The efficiency of protein compartmentalization into the secretory pathway. *Mol. Biol. Cell* **16**, 279–291 (2005).
7. Kim, S. J. & Hegde, R. S. Cotranslational partitioning of nascent prion protein into multiple populations at the translocation channel. *Mol. Biol. Cell* **13**, 3775–3786 (2002).
8. Rane, N. S., Chakrabarti, O., Feigenbaum, L. & Hegde, R. S. Signal sequence insufficiency contributes to neurodegeneration caused by transmembrane prion protein. *J. Cell Biol.* **188**, 515–526 (2010).
9. Orsi, A., Fioriti, L., Chiesa, R. & Sitia, R. Conditions of endoplasmic reticulum stress favor the accumulation of cytosolic prion protein. *J. Biol. Chem.* **281**, 30431–30438 (2006).
10. Drisaldi, B. *et al.* Mutant PrP is delayed in its exit from the endoplasmic reticulum, but neither wild-type nor mutant PrP undergoes retrotranslocation prior to proteasomal degradation. *J. Biol. Chem.* **278**, 21732–21743 (2003).
11. Ma, J. & Lindquist, S. Conversion of PrP to a self-perpetuating PrPSc-like conformation in the cytosol. *Science* **298**, 1785–1788 (2002).
12. Chakrabarti, O. & Hegde, R. S. Functional depletion of mahogunin by cytosolically exposed prion protein contributes to neurodegeneration. *Cell* **137**, 1136–1147 (2009).
13. Ma, J., Wollmann, R. & Lindquist, S. Neurotoxicity and neurodegeneration when PrP accumulates in the cytosol. *Science* **298**, 1781–1785 (2002).
14. Rane, N. S., Kang, S. W., Chakrabarti, O., Feigenbaum, L. & Hegde, R. S. Reduced translocation of nascent prion protein during ER stress contributes to neurodegeneration. *Dev. Cell* **15**, 359–370 (2008).
15. Buchberger, A., Bukau, B. & Sommer, T. Protein quality control in the cytosol and the endoplasmic reticulum: brothers in arms. *Mol. Cell* **40**, 238–252 (2010).
16. McDonough, H. & Patterson, C. CHIP: a link between the chaperone and proteasome systems. *Cell Stress Chaperones* **8**, 303–308 (2003).
17. Leznicki, P., Clancy, A., Schwappach, B. & High, S. Bat3 promotes the membrane insertion of tail-anchored proteins. *J. Cell Sci.* **123**, 2170–2178 (2010).
18. Berndt, U., Oellerer, S., Zhang, Y., Johnson, A. E. & Rospert, S. A signal-anchor sequence stimulates signal recognition particle binding to ribosomes from inside the exit tunnel. *Proc. Natl Acad. Sci. USA* **106**, 1398–1403 (2009).
19. Keenan, R. J., Freymann, D. M., Stroud, R. M. & Walter, P. The signal recognition particle. *Annu. Rev. Biochem.* **70**, 755–775 (2001).
20. Stefanovic, S. & Hegde, R. S. Identification of a targeting factor for posttranslational membrane protein insertion into the ER. *Cell* **128**, 1147–1159 (2007).
21. Lehner, B. *et al.* Analysis of a high-throughput yeast two-hybrid system and its use to predict the function of intracellular proteins encoded within the human MHC class III region. *Genomics* **83**, 153–167 (2004).
22. Park, S. H. *et al.* The cytoplasmic Hsp70 chaperone machinery subjects misfolded and endoplasmic reticulum import-incompetent proteins to degradation via the ubiquitin-proteasome system. *Mol. Biol. Cell* **18**, 153–165 (2007).
23. Eisele, F. & Wolf, D. H. Degradation of misfolded protein in the cytoplasm is mediated by the ubiquitin ligase Ubr1. *FEBS Lett.* **582**, 4143–4146 (2008).
24. Heck, J. W., Cheung, S. K. & Hampton, R. Y. Cytoplasmic protein quality control degradation mediated by parallel actions of the E3 ubiquitin ligases Ubr1 and San1. *Proc. Natl Acad. Sci. USA* **107**, 1106–1111 (2010).
25. Nillegoda, N. B. *et al.* Ubr1 and Ubr2 function in a quality control pathway for degradation of unfolded cytosolic proteins. *Mol. Biol. Cell* **21**, 2102–2116 (2010).
26. Minami, R. *et al.* BAG-6 is essential for selective elimination of defective proteasomal substrates. *J. Cell Biol.* **190**, 637–650 (2010).
27. Ernst, R. *et al.* Enzymatic blockade of the ubiquitin-proteasome pathway. *PLoS Biol.* **8**, e1000605 (2011).
28. Wang, Q. *et al.* A chaperone holdase maintains polypeptides in soluble states for proteasome degradation. *Mol. Cell* doi:10.1016/j.molcel.2011.05.010 (in the press).
29. Garrison, J. L., Kunkel, E. J., Hegde, R. S. & Taunton, J. A substrate-specific inhibitor of protein translocation into the endoplasmic reticulum. *Nature* **436**, 285–289 (2005).
30. Sharma, A., Mariappan, M., Appathurai, S. & Hegde, R. S. *In vitro* dissection of protein translocation into the mammalian endoplasmic reticulum. *Methods Mol. Biol.* **619**, 339–363 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to E. Whiteman and X. Li for carrying out the initial experiments for parts of this project, S.W. Kang, S. Shao, and Z. Zhang for discussions, P. Sengupta, J. Magadán, and C. Ott for constructs, J. Taunton and J. Garrison for cotransin, S. Shao for comments on the manuscript, and Y. Ye for discussions and sharing results before publication. This work was supported by the Intramural Research Program of the National Institutes of Health (R.S.H.) and a postdoctoral fellowship from The Wenner-Gren Foundations (T.H.).

**Author Contributions** T.H. performed most of the experiments, with contributions from A.S. (ubiquitination assays in modified lysates), M.M. (defining the substrate specificity of BAG6), H.D.E. (characterizing the Fr-RRL system), E.G. (BAG6 crosslinking analysis) and R.S.H. (*in vivo* studies). R.S.H. conceived the project, guided the experiments and wrote the paper with input from all of the authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to R.S.H. ([hegde.science@gmail.com](mailto:hegde.science@gmail.com)).



## METHODS

**Plasmids and antibodies.** The SP64 vector-based constructs encoding bovine preprolactin, PrP, ΔSS-PrP (lacking residues 2–22), ΔSSAGPI-PrP (additionally lacking residues 232–254) and HA-tagged PrP (with the epitope inserted at codon 50) have been characterized previously<sup>3,5,29–32</sup>. PrP-PrP and NPY-PrP encode versions in which the N-terminal signal sequence (residues 1–22) of PrP was replaced<sup>5</sup> with that of either bovine preprolactin or human neuropeptide Y (NPY). N3a-PrP contains a mutated signal sequence (WL was replaced with DD at residues 7 and 8) that is translocation deficient<sup>5</sup>. The lysine-free version of PrP was provided by C. Ott and made by standard mutagenesis methods. Wild-type Sec61β (appended at the C terminus with an epitope recognized by the 3F4 antibody), Sec61β(3R), Sec61β-CFP and CFP–Sec61β have been described previously<sup>4,20</sup>. Sec61β-TR (referred to as TR-β in the text and figures) contains the TMD of the human transferrin receptor (IAVIVFLLIGFMIGYLYG) at codon 50 in the cytosolic domain of Sec61β<sup>4</sup>. This positions the TMD of TR outside the ribosomal tunnel when the Sec61β TMD is inside the tunnel<sup>4</sup>. RT-β contains an irrelevant hydrophilic sequence (YKPYIMNPIKKTTITAI) at the same position<sup>4</sup>. GFP, SS/GPI-GFP (containing the N-terminal signal sequence of bovine preprolactin and the C-terminal GPI anchoring sequence of PrP), ManII-GFP (containing the N-terminal type II signal anchor domain of Golgi α-mannosidase II) and SiT-GFP (containing the type II signal anchor domain of sialyl transferase) have been described previously<sup>32–34</sup>. The plasmid encoding Vpu (a type-I-signal-anchored membrane protein from HIV-1) was obtained from J. Bonifacino and J. Magadán<sup>35</sup>. An expression plasmid for bovine rhodopsin has been characterized<sup>29</sup>. For translations of full-length products, the open reading frames were PCR amplified using a forward 5′ primer annealing to or encoding an SP6 or T7 promoter, and a reverse primer in the 3′ untranslated region at least 100 nucleotides beyond the stop codon. For RNCs, the reverse primer annealed in the coding region and lacked a stop codon. PrP and Vpu RNCs included the entire open reading frame except for the stop codon. The RNCs of β-CFP encoded 46 residues beyond the TMD such that this domain would fully emerge from the ribosome. Similarly, the RNCs of TR-β and RT-β encoded up to and including the TMD of Sec61β such that the TR and RT sequences emerge from the ribosome. Genetic constructs encoding BAG6-Flag and ΔUBL-BAG6-Flag (lacking residues 15–89 of BAG6)—both encoding human BAG6 containing a C-terminal Flag epitope—were subcloned into a mammalian expression vector by using standard methods. Expression vectors for human TRC35 and UBL4A containing C-terminal Flag tags were obtained from OriGene. Expression vectors for shRNAs directed against human BAG6 were from OriGene. The target sequences were TGACGGCT CTGCTGTGGATGTTACATCA and CAGCTATGTCATGGTTGGAACCT TCAATC. The irrelevant sequence used as a control was GCACTACCAG AGCTAACTCAGATAGTACT. Antibodies specific for BAG6, TRC40, TRC35, UBL4A and Sec61β have been described previously<sup>4,36</sup>. Anti-SRP54 (BD Biosciences), anti-ubiquitin (BIOMOL), and 3F4 anti-PrP monoclonal antibodies (Signet) were purchased.

**In vitro translation.** *In vitro* transcription and translation in RRL was carried out with minor modifications to published procedures<sup>30</sup>. The most notable change was the inclusion in most experiments of 10 μM His-tagged ubiquitin (Boston Biochem) to facilitate the subsequent isolation of ubiquitinated products. Preliminary experiments showed that, at this concentration, endogenous ubiquitin was more than 90% competed out, resulting in few or no untagged ubiquitinated products. Translation times, unless otherwise indicated, were 1 h at 32 °C. Shorter times for tail-anchored proteins (as used in our earlier studies) resulted in very little ubiquitination<sup>4,20</sup>, presumably because saturation of TRC40 is required before substrates occupy the Bag6 complex<sup>4</sup>. To generate RNCs, the translation times were typically reduced to 30 min to minimize spontaneous release or hydrolysis of the tRNA. Translocation assays into rough microsomes<sup>5</sup>, inhibition by cotransin<sup>29</sup> and inactivation with NEM<sup>37</sup> treatment were carried out as previously described. For direct analysis or downstream immunoprecipitation, translation reactions were stopped, and the proteins were denatured using 1% SDS and heating to 100 °C. For other applications requiring native complexes (for example, crosslinking, affinity purification or downstream assays), samples were placed on ice, and subsequent manipulations were performed at 0–4 °C.

**Sucrose gradient separation and crosslinking.** To generate RNCs, translation reactions (typically 200 μl volume) were chilled on ice and immediately layered onto 2-ml 10–50% sucrose gradients in physiological salt buffer (PSB; 100 mM potassium acetate, 50 mM HEPES, pH 7.4, and 2 mM magnesium acetate). Centrifugation was carried out for 1 h at 55,000 r.p.m. at 4 °C in a TLS-55 rotor (Beckman), after which 200 μl fractions were removed from the top. The peak ribosomal fractions (6 and 7) were pooled and used as the RNCs. These were used immediately or flash frozen in liquid nitrogen for later use in RNC crosslinking or ubiquitination experiments. Chemical crosslinking experiments were essentially carried out as described previously<sup>4,20</sup>. Chilled translation reactions were layered

onto 2-ml 5–25% sucrose gradients in PSB and centrifuged for 5 h at 55,000 r.p.m. at 4 °C in a TLS-55 rotor, after which 200 μl fractions were removed from the top. Crosslinking experiments used 250 μM BMH, except for in experiments to detect SRP interaction, which used 200 μM DSS. Reactions were carried out for 30 min at either 0 °C (BMH) or 25 °C (DSS) and quenched with 25 mM 2-mercaptoethanol (BMH) or 100 mM Tris (DSS). The samples were subsequently denatured and subjected to direct analysis or immunoprecipitation as described below. Photocrosslinking was carried out by following published methods<sup>38</sup>, except that we used the Fr-RRL system for translation and benzophenone-modified lysyl-tRNA (tRNA Probes). The absence of endogenous charged tRNAs and haemoglobin increased photocrosslinker incorporation and photolysis, respectively. Photolysis was carried out for 15 min on ice, and the samples were analysed directly.

**Modified translation extracts.** Fr-RRL was typically prepared from 25 ml RRL (Green Hectares) that had first been treated with haemin and micrococcal nuclease. Its characterization will be described in a future publication, but its preparation is as follows. All procedures were carried out on ice or at 4 °C. The lysate was centrifuged at 100,000 r.p.m. for 40 min in a TLA100.4 rotor (Beckman). The supernatants were pooled, and the tubes rinsed (without disrupting the ribosomal pellet) with an equal volume of column buffer (20 mM Tris, pH 7.5, 20 mM KCl, 0.1 mM EDTA and 10% glycerol), which was added to the supernatant. The pellet was resuspended by dounce homogenization in ribosome wash buffer (RWB; 20 mM HEPES, pH 7.5, 100 mM potassium acetate, 1.5 mM magnesium acetate and 0.1 mM EDTA), layered onto a 1 M sucrose cushion in RWB, and re-isolated by centrifugation at 100,000 r.p.m. for 1 h in a TLA100.4 rotor. The final pellet was resuspended in one-tenth of the original lysate volume and defined as 'native ribosomes'. The ribosome-free supernatant from above was applied to a 10 ml DEAE column at a flow rate of ~1 ml min<sup>-1</sup> and washed with column buffer until the red haemoglobin was removed (~50 ml). The elution was carried out in a single step with 50 ml column buffer containing 300 mM KCl. The eluate was adjusted slowly with solid ammonium sulphate to 75% saturation (at 4 °C) with constant stirring. After 1 h mixing, the precipitate was recovered by centrifugation at 15,000 r.p.m. in a JA-17 rotor (Beckman). The supernatant was discarded, and the pellet was dissolved in a minimal volume (~8 ml) of dialysis buffer (20 mM HEPES, pH 7.4, 100 mM potassium acetate, 1.5 mM magnesium acetate, 10% glycerol and 1 mM DTT). This solution was dialysed against two changes of dialysis buffer overnight, recovered, adjusted to 10–12 ml (that is, twice the original concentration) and flash frozen in liquid nitrogen. To make a translation-competent Fr-RRL, the native ribosomes and dialysed DEAE eluate were adjusted to 72 mM potassium acetate, 2.5 mM magnesium acetate, 10 mM HEPES, pH 7.4, 2 mM DTT, 0.2 mg ml<sup>-1</sup> liver tRNA, 1 mM ATP, 1 mM GTP, 12 mM creatine phosphate, 40 μg ml<sup>-1</sup> creatine kinase, 40 μM each amino acid (except for methionine) and 1 μCi ml<sup>-1</sup> [<sup>35</sup>S]methionine. The concentration of ribosomes and lysate was the same as that for RRL. Immunodepletions of RRL were carried out as described previously<sup>4</sup>.

**Ubiquitination assays.** The human E1 enzyme and all mammalian E2 enzymes were obtained from Boston Biochem. For full-length proteins, translations containing 10 μM His-ubiquitin were carried out for 1 h at 32 °C. In Fr-RRL, post-translational ubiquitination was initiated by adding E2 enzyme to a final concentration of 250 nM and further incubating for 1 h. For RNCs, samples were supplemented (as indicated in the figures) with E1 enzyme (85 nM), E2 enzyme (usually 250 or 500 nM), cytosol (RRL or Fr-RRL, at the same concentration as in the translations), 10 μM His-ubiquitin, an ATP-regenerating system (1 mM ATP, 10 mM creatine phosphate and 40 μg ml<sup>-1</sup> creatine kinase) and 1 mM puromycin. Reaction conditions were 100 mM potassium acetate, 50 mM HEPES, pH 7.4, 5 mM MgCl<sub>2</sub> and 1 mM DTT. Incubation was for 1 h at 32 °C. On-bead ubiquitination of affinity-purified products was carried out under the same conditions, except for without puromycin. To prepare the affinity-purified substrate, translation reactions in Fr-RRL were chilled on ice, diluted to 1 ml in PSB and incubated with immobilized antibodies against the HA epitope (for PrP-HA and Vpu-HA) or Sec61β. In Supplementary Fig. 9, the translation reactions were supplemented with Flag-tagged BAG6 or ΔUBL-BAG6 (each added to twofold excess above endogenous BAG6 levels), and anti-Flag beads (Sigma) were used for the pull-down. After 1 h, the resin was washed five times in PSB, and the residual buffer was carefully removed before adding the ubiquitination components as above. The reaction was incubated with constant low-level shaking (in a Thermomixer, Eppendorf) at 32 °C for 1 h. SDS (1%) was added directly to the reactions, which were analysed directly and after ubiquitin pull-downs.

**Cell culture studies.** Culture, transfection and immunoblotting analysis of N2a cells (dominant-negative inhibition experiments) and HeLa cells (for shRNA experiments) were carried out as described previously<sup>2,3</sup>. Cells were seeded in 24-well dishes the day before transfection. For the dominant-negative experiments, the plasmids were mixed in the ratios indicated in Supplementary Fig. 20 and transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. At 24 h after transfection, the cells were harvested in 1% SDS;



the DNA was sheared by vortexing and boiling; and the total sample was analysed by SDS-PAGE and immunoblotting. For shRNA experiments, each well received a mixture of 550 ng shRNA plasmid, 200 ng PrP expression plasmid and 50 ng CFP expression plasmid. Transfection was effected with Lipofectamine 2000. Examination of CFP fluorescence verified at least 50% transfection efficiency. The cells were cultured for ~100 h before collection and analysis by immunoblotting.

**BAG6 purification.** Full-length BAG6 or ΔUBL-BAG6 tagged at the C terminus with a Flag epitope was overexpressed by transient transfection of HEK-293T cells. TransIT reagent (Mirus) was used. After 3 days of expression, the cells were collected in 50 mM HEPES, pH 7.4, 150 mM potassium acetate, 5 mM magnesium acetate and 1% deoxy Big CHAP. The soluble extract was incubated with immobilized anti-Flag antibodies (Sigma) with constant mixing, and the resin was washed four times with high salt lysis buffer containing 400 mM potassium acetate and then twice with detergent-free lysis buffer containing 230 mM potassium acetate. Elution was carried out with 1 mg ml<sup>-1</sup> competing peptide at room temperature. The final protein was checked by using colloidal Coomassie blue (Supplementary Fig. 16), and its concentration relative to that in RRL was determined by immunoblotting of serial dilutions. Blotting also confirmed the lack of TRC35 and UBL4A in BAG6 prepared by this method.

**Miscellaneous biochemistry.** Immunoprecipitation assays were carried out as described previously<sup>5,36</sup>. Pull-down assays with Co<sup>2+</sup> immobilized on chelating sepharose were performed on samples denatured in boiling 1% SDS and then diluted tenfold in cold (4 °C) 0.5% Triton X-100, 25 mM HEPES, 100 mM NaCl and 10 mM imidazole. The complete denaturation step is essential for samples containing RRL because the haemoglobin is a strong Co<sup>2+</sup>-binding protein in its

native state. Typically, 10 μl packed resin was used per sample, and after incubation for 1–2 h at 4 °C, the resin was washed three times in the above buffer and eluted in SDS-PAGE sample buffer containing 20 mM EDTA. SDS-PAGE was carried out using 8.5% or 12% tricine gels. Figures were prepared using the programs Photoshop and Illustrator (Adobe).

31. Emerman, A. B., Zhang, Z. R., Chakrabarti, O. & Hegde, R. S. Compartment-restricted biotinylation reveals novel features of prion protein metabolism *in vivo*. *Mol. Biol. Cell* **21**, 4325–4337 (2010).
32. Ashok, A. & Hegde, R. S. Retrotranslocation of prion proteins from the endoplasmic reticulum by preventing GPI signal transamidation. *Mol. Biol. Cell* **19**, 3463–3476 (2008).
33. Cole, N. B. *et al.* Diffusional mobility of Golgi proteins in membranes of living cells. *Science* **273**, 797–801 (1996).
34. Wu, M. M. *et al.* Organelle pH studies using targeted avidin and fluorescein-biotin. *Chem. Biol.* **7**, 197–209 (2000).
35. Magadán, J. G. *et al.* Multilayered mechanism of CD4 downregulation by HIV-1 Vpu involving distinct ER retention and ERAD targeting steps. *PLoS Pathogens* **6**, e1000869 (2010).
36. Fons, R. D., Bogert, B. A. & Hegde, R. S. Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. *J. Cell Biol.* **160**, 529–539 (2003).
37. Gilmore, R., Blobel, G. & Walter, P. Protein translocation across the endoplasmic reticulum. I. Detection in the microsomal membrane of a receptor for the signal recognition particle. *J. Cell Biol.* **95**, 463–469 (1982).
38. Krieg, U. C., Walter, P. & Johnson, A. E. Photocrosslinking of the signal sequence of nascent prolactin to the 54-kilodalton polypeptide of the signal recognition particle. *Proc. Natl Acad. Sci. USA* **83**, 8604–8608 (1986).

# The ELF4–ELF3–LUX complex links the circadian clock to diurnal control of hypocotyl growth

Dmitri A. Nusinow<sup>1,2</sup>, Anne Helfer<sup>1,2</sup>, Elizabeth E. Hamilton<sup>1,2</sup>, Jasmine J. King<sup>1,2</sup>, Takato Imaizumi<sup>1†</sup>, Thomas F. Schultz<sup>1†</sup>, Eva M. Farré<sup>1†</sup> & Steve A. Kay<sup>1,2</sup>

The circadian clock is required for adaptive responses to daily and seasonal changes in environmental conditions<sup>1–3</sup>. Light and the circadian clock interact to consolidate the phase of hypocotyl cell elongation to peak at dawn under diurnal cycles in *Arabidopsis thaliana*<sup>4–7</sup>. Here we identify a protein complex (called the evening complex)—composed of the proteins encoded by *EARLY FLOWERING 3* (*ELF3*), *ELF4* and the transcription-factor-encoding gene *LUX ARRHYTHMO* (*LUX*; also known as *PHYTOCLOCK 1*)—that directly regulates plant growth<sup>8–12</sup>. *ELF3* is both necessary and sufficient to form a complex between *ELF4* and *LUX*, and the complex is diurnally regulated, peaking at dusk. *ELF3*, *ELF4* and *LUX* are required for the proper expression of the growth-promoting transcription factors encoded by *PHYTOCHROME INTERACTING FACTOR 4* (*PIF4*) and *PIF5* (also known as *PHYTOCHROME INTERACTING FACTOR 3-LIKE 6*) under diurnal conditions<sup>4,6,13</sup>. *LUX* targets the complex to the promoters of *PIF4* and *PIF5* *in vivo*. Mutations in *PIF4* and/or *PIF5* are epistatic to the loss of the *ELF4*–*ELF3*–*LUX* complex, suggesting that regulation of *PIF4* and *PIF5* is a crucial function of the complex. Therefore, the evening complex underlies the molecular basis for circadian gating of hypocotyl growth in the early evening.

The circadian clock is an endogenous molecular oscillator with a period of ~24 h that is almost ubiquitous<sup>1</sup>. In plants, multiple interlocking transcriptional feedback loops contribute to the robust architecture of this oscillator network<sup>3</sup>. The clock functions to enable anticipation of diurnal, rhythmic environmental changes, allowing optimal phasing of molecular, physiological and behavioural responses to specific times of day<sup>2</sup>. Plant growth is a physiological response that is controlled by both the clock and the changes in light conditions; under diurnal growth conditions, maximal plant growth occurs at the end of night<sup>4–7</sup>.

*ELF3* and *ELF4* were first identified in genetic screens for photoperiodism mutants and were found to regulate circadian rhythms<sup>8–10,14</sup>. *ELF3* and *ELF4* encode plant-specific nuclear proteins with no known functional domains<sup>9,10,15,16</sup>. *LUX* is a single-MYB-domain-containing, SHAQYF-type GARP transcription factor that was identified in a genetic screen for long hypocotyl mutants and aberrant circadian-regulated gene expression<sup>11,12</sup>. The mutants *elf3*, *elf4* and *lux* share multiple phenotypes, including an arrhythmic circadian oscillator, abnormal hypocotyl growth in diurnal cycles, and early flowering<sup>4,8–12,14,15,17</sup>. *ELF3*, *ELF4* and *LUX* showed similar expression profiles in microarray experiments (Supplementary Fig. 1; DIURNAL database, <http://diurnal.cgrb.oregonstate.edu> refs 18, 19), and these expression profiles were confirmed by quantitative PCR with reverse transcription (RT–PCR) analysis under both diurnal and circadian conditions (Fig. 1a).

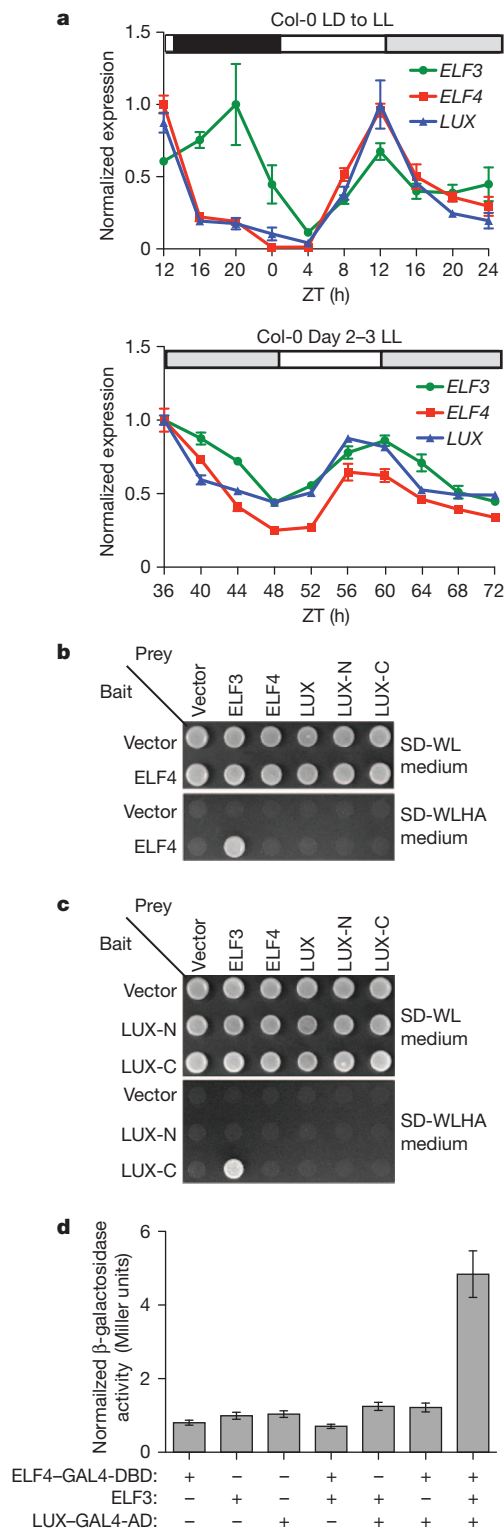
The similarities in expression patterns and phenotypes prompted us to test whether these proteins could interact. Using a yeast two-hybrid

assay, we found that *ELF4* interacted with *ELF3* (Fig. 1b). In addition, when *LUX* fragments were used as baits (full-length *LUX* showed auto-activation, data not shown), *ELF3* showed an interaction with *LUX*-C (amino acids 144–324), which contains the DNA-binding domain of *LUX*<sup>11,12,20</sup>, but not with *LUX*-N (amino acids 1–143) (Fig. 1c). *ELF4* did not interact with *LUX* or either *LUX* fragment (Fig. 1b, c). As *ELF3* could interact independently with either *ELF4* or *LUX*, we proposed that *ELF3* might form a complex between these two proteins. To test this, *ELF3* was used in a yeast three-hybrid system in combination with the fusion proteins *ELF4*–GAL4–DNA binding domain (GAL4–DBD) and/or *LUX*–GAL4–activation domain (GAL4–AD). Activation of the reporter was observed only when all three proteins were present, suggesting that *ELF3* was sufficient to bridge an interaction between *ELF4* and *LUX* (Fig. 1d).

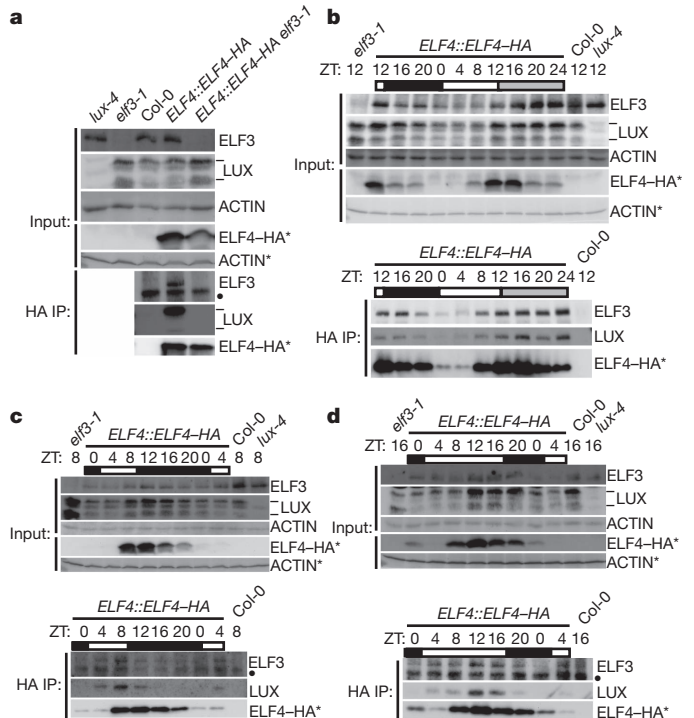
Next, we tested whether *ELF4*, *ELF3* and *LUX* interact *in vivo*. Antibodies were developed against *ELF3* and *LUX*, and an *ELF4::ELF4*–HA construct was introduced into the *elf4-2* mutant<sup>21</sup>. The encoded haemagglutinin (HA)-tagged *ELF4* protein is probably functional, because we identified transformants that rescued the hypocotyl length (Supplementary Fig. 2a) and circadian *CHLOROPHYLL A/B-BINDING PROTEIN::LUCIFERASE* (*CAB2::LUC*) rhythmicity, albeit with a shorter period than that of the wild type (Supplementary Fig. 2b–d). We then asked whether *ELF4*–HA could co-immunoprecipitate endogenous *ELF3* and/or *LUX* at Zeitgeber time 12 (ZT12) (Fig. 2a). We found that *ELF4*–HA could co-immunoprecipitate both *ELF3* and *LUX* (Fig. 2a and Supplementary Fig. 2f). The experiments in yeast suggested that *ELF3* bridges an interaction between *ELF4* and *LUX* and that *ELF3* would be necessary for the co-immunoprecipitation of *LUX* by *ELF4*–HA. To test this, we introduced *elf3-1* into the *ELF4::ELF4*–HA *elf4-2* transgenic line and immunoprecipitated *ELF4*–HA. Although similar amounts of *ELF4* and *LUX* were present in the extracts, *LUX* did not co-immunoprecipitate with *ELF4*–HA (Fig. 2a). These results show that *ELF3* is necessary for *in vivo* formation of the tripartite complex that includes *ELF4* and *LUX*. Furthermore, hypocotyl length in *elf3-1 elf4-3* and *elf3-1 lux-4* double mutants grown under a 12 h light and 12 h dark (12L:12D) cycle did not show additive effects over *elf3* (Supplementary Fig. 3). These results are consistent with the hypothesis that *ELF3*, *ELF4* and *LUX* function together as a complex to regulate common pathways.

Because *ELF4*, *ELF3* and *LUX* messenger RNA levels oscillate and peak with a similar phase, we analysed the dynamics of the protein levels under diurnal cycles. Tissue from the *ELF4::ELF4*–HA transgenic line was harvested every 4 h, starting at ZT12, under 12L:12D cycles, and then after transfer to constant light at ZT0 the following day. *ELF3*, *LUX* and *ELF4*–HA protein levels peaked at ZT12, declined during the night, reached a trough between ZT0 and ZT4 and then increased again (Fig. 2b and Supplementary Fig. 4). The levels of all three proteins remained elevated into the subjective dark period relative to their

<sup>1</sup>Section of Cell & Developmental Biology, Division of Biological Sciences, University of California San Diego 9500 Gilman Drive, La Jolla, California 92093-0130, USA. <sup>2</sup>Center for Chronobiology, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0116, USA. †Present addresses: Department of Biology, University of Washington, 24 Kincaid Hall, Box 351800, Seattle, Washington 98195-1800, USA (T.I.); Nicholas School of the Environment, Duke University Marine Laboratory, 135 Duke Marine Lab Road, Beaufort, North Carolina 28516, USA (T.F.S.); Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824-1312, USA (E.M.F.).



**Figure 1** | *ELF3*, *ELF4* and *LUX* are co-expressed, and *ELF3* directly interacts with both *ELF4* and *LUX* in yeast. **a**, Expression analysis by RT-PCR of *ELF3*, *ELF4* and *LUX* under diurnal or circadian conditions. Normalization is relative to the maximum. The rectangles above the graphs represent the light conditions during harvesting: black, lights off; white, lights on; and grey, lights on during subjective night. Error bars, s.e.m.;  $n = 3$ . LD, 12L:12D; LL, constant light. **b**, Yeast two-hybrid assay between *ELF4* and each of *ELF3*, *ELF4*, *LUX*, *LUX-N* and *LUX-C*. These experiments were repeated twice. **c**, Yeast two-hybrid assay carried out as for **b**, between *LUX* and each of *ELF3*, *ELF4* and *LUX*. **d**, Yeast three-hybrid assay assessing combinations of *ELF4*-GAL4-DBD, *LUX*-GAL4-AD and *ELF3*. Data are presented as fold induction over control vectors. Error bars, s.e.m.;  $n = 4$ .



**Figure 2** | *ELF3* bridges a diurnally regulated complex containing *ELF4* and *LUX* *in vivo*. **a**, *ELF3* is necessary for *ELF4* and *LUX* to co-precipitate *in vivo*. Immunoprecipitations (IPs) were performed on day 12 at ZT12 of a 12L:12D cycle. **b**, *ELF3*, *ELF4* and *LUX* oscillate over time and form a complex. **c**, **d**, EC formation in short and long days. Seedlings were grown under short-day or long-day photoperiods (8L:16D cycle or 16L:8D cycle, respectively) and harvested beginning at ZT0 on day 12. Experiments were performed three times with similar results. **a–d**, Rectangles indicate light conditions during harvesting as denoted in Fig. 1a. Upper panels show inputs, and lower panels HA IPs. —, each of the two *LUX* isoforms (low and high molecular weight); \*, *ELF4* separated in a 15% gel; •, background arising from HA-crosslinked beads.

respective time points in the dark, and the protein peak was shifted from ZT12 to the middle of the subjective night (Fig. 2b). Comparable results were observed for *ELF3* and *LUX* levels in wild-type seedlings (Supplementary Fig. 5). To assay time-dependent formation of the *ELF4*-*ELF3*-*LUX* complex (denoted the evening complex, EC), *ELF4*-HA was immunoprecipitated from the diurnal samples. The formation of the EC followed the same pattern as that of its composite parts, suggesting that these proteins would associate when present (Fig. 2b).

Photoperiodic control of flowering and growth is compromised in *elf3*, *elf4* and *lux* mutants<sup>4,8–12,14,15,17,22</sup>. To determine how *ELF4*, *ELF3* and *LUX* respond to altered photoperiods, we analysed the levels and formation of the EC in plants grown under short days (8L:16D) and long days (16L:8D). Peak levels of *ELF4*, *ELF3* and *LUX* followed their respective mRNA profiles under different photoperiods (Fig. 2c, d and Supplementary Fig. 4), similar to the findings of previous reports<sup>9,10,15,22</sup>. EC formation was also sensitive to photoperiod, peaking earlier in short days than in long days (Fig. 2c, d).

To investigate the molecular role of the EC, we focused on the diurnal hypocotyl growth phenotype shared by all mutants<sup>5,8–11,20</sup>. Previous work demonstrated that the basic helix-loop-helix transcription factors *PIF4* and *PIF5* are crucial for determining the hypocotyl elongation rate in seedlings, and that the genes encoding both factors act downstream of light- and clock-signalling pathways<sup>4,6,7,13,23</sup>. Expression of *PIF4* and *PIF5* was nearly antiphasic to that of the EC under different photocycles (Supplementary Fig. 6). This raised the possibility that the EC may be repressing the transcription of *PIF4* and *PIF5*, which is consistent with recent reports that *ELF3* and *LUX* act as

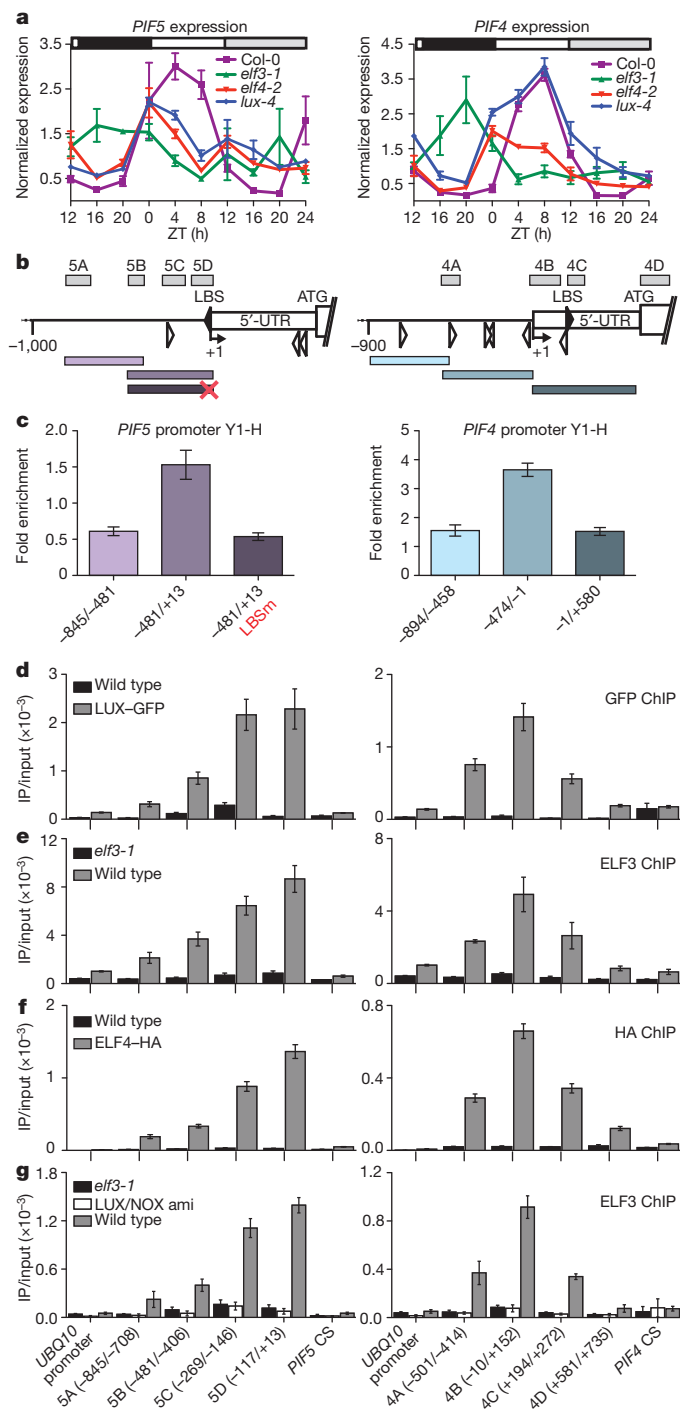


transcriptional repressors in the circadian clock<sup>20,24</sup>. The levels of *PIF4* and *PIF5* mRNA are elevated in *elf3-1*, *elf4-2* and *lux-4* mutants compared with the wild type, particularly during the early evening (Fig. 3a). Recent work demonstrated that the addition of an activation domain to LUX (LUX-VP64) induced a neomorphic hypocotyl elongation phenotype<sup>20</sup>, and we found that *PIF4* and *PIF5* expression levels were increased in this background (Supplementary Fig. 7). These results, as well as the presence of full consensus LUX-binding sites (LBSs)<sup>20</sup> in the 5'-untranslated region of both *PIF4* and *PIF5* (Fig. 3b), suggested that LUX may participate directly in the modulation of *PIF4* and *PIF5* expression. Indeed, LUX was able to directly bind to the *PIF4* and *PIF5* promoters in yeast, and LUX binding to the *PIF5* promoter (from -481 to +13 base pairs) was lost when the consensus LBS was mutated (Fig. 3c).

To determine whether components of the EC were bound to the *PIF4* and *PIF5* promoters *in vivo*, chromatin immunoprecipitation (ChIP) assays were performed in *LUX::LUX-GFP* transgenic lines and then the *PIF4* and *PIF5* promoter sequences were amplified. These experiments revealed *in vivo* binding to the LBS in the promoters of *PIF4* and *PIF5* but not to control sequences in the coding regions of these genes or in the *POLYUBIQUITIN10* (*UBQ10*) promoter (Fig. 3d). The formation of the EC (Fig. 2) suggested that all of its components might participate in the regulation of *PIF4* and *PIF5* expression; therefore, we performed similar ChIP experiments for *ELF3* and *ELF4-HA*. We found that *ELF3* and *ELF4-HA* showed specific enrichment at the *PIF4* and *PIF5* promoter sequences that were bound by LUX (Fig. 3e, f). Additionally, *ELF3* ChIP experiments performed at the trough of EC levels (ZT2) showed a lower specific enrichment than those performed at ZT14 (Supplementary Fig. 8).

The localization pattern of the EC components on the *PIF4* and *PIF5* promoters suggested that the transcription factor LUX might be responsible for recruitment. *ELF3* ChIP experiments in *lux-4* seedlings demonstrated that less *ELF3* was recruited to the *PIF4* and *PIF5* promoters in these mutants but that recruitment was not completely abrogated (Supplementary Fig. 9). Previous work identified a MYB-domain-containing transcription factor highly similar to LUX, named NOX (At5g59570)<sup>11,12,20</sup>. NOX binds sequences that are similar to those bound by LUX in yeast<sup>20</sup> and was also able to form a complex with *ELF4* and *ELF3* (Supplementary Fig. 10a). We designed an artificial microRNA (amiRNA) using a web-based amiRNA algorithm (<http://wmd3.weigelworld.org/cgi-bin/webapp.cgi>) and generated an amiRNA-transgenic line in which the levels of both NOX and LUX would simultaneously be reduced (denoted LUX/NOX ami)<sup>25,26</sup>. LUX protein and NOX expression levels were reduced in this line (Supplementary Fig. 10b, c), which showed similar defects in circadian rhythms to *lux-4* mutants (Supplementary Fig. 10e, f); however, we observed an increase in hypocotyl length and *PIF4* and *PIF5* expression level compared with *lux-4* mutants (Supplementary Fig. 10d, g). When *ELF3* ChIP assays were performed in the LUX/NOX ami line, we observed a loss of the *ELF3* signal at the *PIF4* and *PIF5* promoters (Fig. 3g). *ELF3* was still present in extracts from these plants (Supplementary Fig. 10c), suggesting that the recruitment of *ELF3* (and therefore the EC) is mediated by both LUX and NOX.

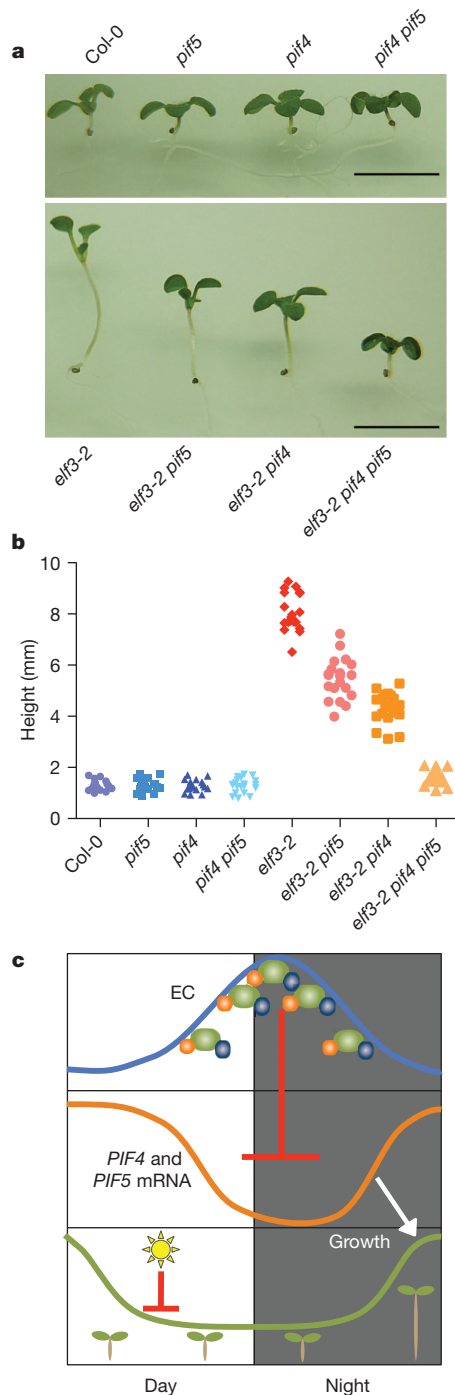
Previous reports showed that ectopic overexpression of the MYB-domain-containing transcription factors encoded by *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*; in the *CCA1-OX* line) or *LATE ELONGATED HYPOCOTYL* (*LHY*; in *lhy-1* mutants) resulted in phenotypes similar to those of *elf3*, *elf4* or *lux-4* (Fig. 3a and Supplementary Fig. 11). As *CCA1* and *LHY* form a complex that controls the expression of evening-element-containing genes<sup>27</sup>, such as *ELF4* and *LUX*<sup>11,28</sup>, the misexpression of *PIF4* and *PIF5* seen in *CCA1-OX* or *lhy-1* lines could be a result of EC misregulation. Therefore, we analysed the expression of *ELF4*, *ELF3* and *LUX* in the *lhy-1* background using the DIURNAL database<sup>18,19</sup>. We found that *ELF4* was clamped low, whereas *ELF3* and *LUX* were shifted 4 h and 12 h later, respectively (Supplementary Fig. 11). These results are consistent with the



**Figure 3 | The EC regulates *PIF5* and *PIF4* expression through recruitment by LUX.** **a**, *PIF5* and *PIF4* expression in *elf3*, *elf4*, *lux* and wild type (Col-0). Rectangles indicate light conditions as denoted in Fig. 1a. Error bars, s.e.m.;  $n = 3$ . **b**, *PIF5* (left) and *PIF4* (right) promoters denoting a degenerate (GATWCK or GATWYG) or consensus (GATWCG) LBS, represented by unfilled or filled arrowheads, respectively. Numbers are relative to the transcriptional start site (+1). Rectangles represent ChIP amplicons (top) and fragments for yeast one-hybrid assays (bottom); the red X denotes a mutated LBS. UTR, untranslated region. **c**, Yeast one-hybrid (Y1-H) with LUX-GFP and *PIF5* and *PIF4* promoter fragments (where / denotes a range). The fold enrichment is relative to controls. LBSm, LBS mutant. **d–g**, ChIP on *PIF5* and *PIF4* at ZT14, under extended light conditions: LUX (**d**), *ELF3* (**e**, **g**) and *ELF4* (**f**). **d**, **e**, Data are presented as mean  $\pm$  s.e.m.;  $n = 3$ . **f**, **g**, Data are presented as mean  $\pm$  s.d. (from two technical replicates measured twice). Experiments were repeated with similar results. CS, coding sequence.

circadian clock having a crucial role in the proper expression and phasing of the EC proteins.

If improper regulation of *PIF4* and *PIF5* underlies the hypocotyl growth defects observed in the EC mutants, then loss of *PIF4* and *PIF5*



**Figure 4 | Hypocotyl growth defects are rescued by loss of *PIF5* and *PIF4* in EC component mutant backgrounds.** **a**, Growth defects in the *elf3-2* background require *PIF4* and *PIF5*. Scale bar, 5 mm. **b**, Scatter plot of hypocotyl measurements from the wild type (Col-0), as well as *elf3-2*, *pif4-101* and *pif5-1* single and compound mutants. This experiment was repeated with similar results. **c**, The model represents the action of the EC on *PIF4* and *PIF5* expression during the early evening, which results in the gating of hypocotyl growth in *A. thaliana* seedlings. The circadian-regulated EC represses *PIF4* and *PIF5* expression in the evening. Throughout the day, post-transcriptional light-mediated degradation of *PIF4* and *PIF5* proteins inhibits growth. Near dawn, the concomitant rise in *PIF4* and *PIF5* mRNA and *PIF4* and *PIF5* protein levels promotes growth (white arrow).

should be epistatic to loss of the EC. To test this, we introduced *pif4* and *pif5* mutant alleles into the *elf3-2* mutant background, because mutating *ELF3* caused dissolution of the EC (Fig. 2a). Loss of *PIF4* or *PIF5* additively mitigated the hypocotyl length defect in *elf3-2* (Fig. 4a, b), indicating that the hypocotyl phenotypes of EC mutants are mainly caused by misexpression of *PIF4* and *PIF5*. In addition, loss of *PIF4* and/or *PIF5* did not restore circadian rhythms in an *elf3* background (Supplementary Fig. 12), consistent with *PIF4* and *PIF5* being clock outputs that do not feed back into the oscillator<sup>4</sup>.

In summary, we have identified a novel multiprotein complex that directly links the circadian clock to diurnal regulation of hypocotyl growth. The ELF4–ELF3–LUX complex is regulated by the clock and by light (Figs 1a and 2b–d) and represses the expression of *PIF4* and *PIF5* in the early evening (Fig. 4c). This process is combined with light-regulated turnover of *PIF4* and *PIF5*, allowing maximum hypocotyl growth at dawn under diurnal conditions<sup>4,13</sup> (Fig. 4c). *ELF3* is necessary and sufficient to bring together *ELF4* and *LUX* to form a complex (Figs 1d and 2a), providing a mechanistic framework for understanding the shared phenotypes of EC component mutants in regulating circadian rhythms, growth and flowering. The role of *ELF3* as an adaptor protein is similar to its previously described capacity to modulate GIGANTEA levels through association with CONSTITUTIVELY PHOTOMORPHOGENIC 1 to regulate flowering and circadian rhythms<sup>22</sup>. The EC is composed of multiple proteins that are known to regulate signalling from the environment<sup>4,5,9–12,14–17,20,22,24,28</sup>; therefore, elucidating EC function will ultimately contribute to understanding how biochemical, physiological and developmental outputs are gated by the clock.

## METHODS SUMMARY

All wild-type, mutant and transgenic lines were in the *A. thaliana* ecotype Columbia-0 (Col-0). All transgenic and mutant lines were brought to homozygosity before use. The procedures for *A. thaliana* husbandry, yeast one-hybrid, two-hybrid and three-hybrid analyses, bioluminescent imaging, immunoprecipitation assays, ChIP assays and hypocotyl measurements have been described previously<sup>20,29,30</sup> and were carried out with modifications detailed in the Methods. In all growth chambers, light was supplied at  $80 \mu\text{mol m}^{-2} \text{s}^{-1}$  by cool-white fluorescent bulbs at 22 °C. For yeast two-hybrid analyses, SD-WL medium was used to select for the presence of both bait and prey vectors, and SD-WLHA medium was used to select for an interaction between the bait and the prey proteins. *IPP2*, *APX3* and *Atlg11910* levels were used to normalize real-time PCR expression analyses, and all primers for quantitative PCR are listed in Supplementary Table 1. The *ELF4::ELF4-HA* construct includes a 580-bp promoter sequence cloned from Col-0 DNA that was amplified using primers listed in Supplementary Table 1. The sequence TATGATATCCTTGCCTACCCA is the target of the LUX/NOX ami. Antibodies were generated in rabbits (Sigma Genosys) against either an ELF3-specific peptide (CSIQEERKRYDSSKP) or a full-length LUX protein fused to glutathione *S*-transferase (GST). Antibodies were affinity purified against the same ELF3-specific peptide using a SulfoLink Immobilization Kit (Thermo Scientific) or a GST–LUX affinity column. All immunoprecipitations were performed with Protein G Dynabeads (Invitrogen). For western blotting, ACTIN served as a loading control. Blots for *ELF4* represent 20% of the total immunoprecipitation sample, because *ELF4* needed to be separated on a different, 15%, gel, owing to its low molecular weight. Hypocotyl measurements were performed on evenly spaced seedlings grown under a 12L:12D cycle and measured on day 10.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 15 October 2010; accepted 6 May 2011.

Published online 13 July 2011.

- Wijnen, H. & Young, M. W. Interplay of circadian clocks and metabolic rhythms. *Annu. Rev. Genet.* **40**, 409–448 (2006).
- Yakir, E., Hilman, D., Harir, Y. & Green, R. M. Regulation of output from the plant circadian clock. *FEBS J.* **274**, 335–345 (2007).
- Harmer, S. L. The circadian system in higher plants. *Annu. Rev. Plant Biol.* **60**, 357–377 (2009).
- Nozue, K. *et al.* Rhythmic growth explained by coincidence between internal and external cues. *Nature* **448**, 358–361 (2007).
- Michael, T. P. *et al.* A morning-specific phytohormone gene expression program underlying rhythmic plant growth. *PLoS Biol.* **6**, e225 (2008).

6. Niwa, Y., Yamashino, T. & Mizuno, T. The circadian clock regulates the photoperiodic response of hypocotyl elongation through a coincidence mechanism in *Arabidopsis thaliana*. *Plant Cell Physiol.* **50**, 838–854 (2009).
7. de Montaigu, A., Tóth, R. & Coupland, G. Plant development goes like clockwork. *Trends Genet.* **26**, 296–306 (2010).
8. Zagotta, M. T. *et al.* The *Arabidopsis* *ELF3* gene regulates vegetative photomorphogenesis and the photoperiodic induction of flowering. *Plant J.* **10**, 691–702 (1996).
9. Hicks, K. A., Albertson, T. M. & Wagner, D. R. *EARLY FLOWERING3* encodes a novel protein that regulates circadian clock function and flowering in *Arabidopsis*. *Plant Cell* **13**, 1281–1292 (2001).
10. Doyle, M. R. *et al.* The *ELF4* gene controls circadian rhythms and flowering time in *Arabidopsis thaliana*. *Nature* **419**, 74–77 (2002).
11. Hazen, S. P. *et al.* *LUX ARRHYTHMO* encodes a Myb domain protein essential for circadian rhythms. *Proc. Natl Acad. Sci. USA* **102**, 10387–10392 (2005).
12. Onai, K. & Ishiura, M. *PHYTOCLOCK 1* encoding a novel GARP protein essential for the *Arabidopsis* circadian clock. *Genes Cells* **10**, 963–972 (2005).
13. Lorrain, S., Allen, T., Duek, P. D., Whitelam, G. C. & Fankhauser, C. Phytochrome-mediated inhibition of shade avoidance involves degradation of growth-promoting bHLH transcription factors. *Plant J.* **53**, 312–323 (2008).
14. Hicks, K. A. *et al.* Conditional circadian dysfunction of the *Arabidopsis* early-flowering 3 mutant. *Science* **274**, 790–792 (1996).
15. Liu, X. L., Covington, M. F., Fankhauser, C., Chory, J. & Wagner, D. R. *ELF3* encodes a circadian clock-regulated nuclear protein that functions in an *Arabidopsis* PHYB signal transduction pathway. *Plant Cell* **13**, 1293–1304 (2001).
16. Khanna, R., Kikis, E. A. & Quail, P. H. *EARLY FLOWERING4* functions in phytochrome B-regulated seedling de-etiolation. *Plant Physiol.* **133**, 1530–1538 (2003).
17. Thines, B. & Harmon, F. G. Ambient temperature response establishes *ELF3* as a required component of the core *Arabidopsis* circadian clock. *Proc. Natl Acad. Sci. USA* **107**, 3257–3262 (2010).
18. Mockler, T. C. *et al.* The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching and promoter analysis. *Cold Spring Harb. Symp. Quant. Biol.* **72**, 353–363 (2007).
19. Michael, T. P. *et al.* Network discovery pipeline elucidates conserved time-of-day-specific *cis*-regulatory modules. *PLoS Genet.* **4**, e14 (2008).
20. Helfer, A. *et al.* *LUX ARRHYTHMO* encodes a nighttime repressor of circadian gene expression in the *Arabidopsis* core clock. *Curr. Biol.* **21**, 126–133 (2011).
21. Hazen, S. P. *et al.* Rapid array mapping of circadian clock and developmental mutations in *Arabidopsis*. *Plant Physiol.* **138**, 990–997 (2005).
22. Yu, J. W. *et al.* COP1 and *ELF3* control circadian function and photoperiodic flowering by regulating GI stability. *Mol. Cell* **32**, 617–630 (2008).
23. de Lucas, M. *et al.* A molecular framework for light and gibberellin control of cell elongation. *Nature* **451**, 480–484 (2008).
24. Dixon, L. E. *et al.* Temporal repression of core circadian genes is mediated through *EARLY FLOWERING 3* in *Arabidopsis*. *Curr. Biol.* **21**, 120–125 (2011).
25. Schwab, R., Ossowski, S., Riester, M., Warthmann, N. & Weigel, D. Highly specific gene silencing by artificial microRNAs in *Arabidopsis*. *Plant Cell* **18**, 1121–1133 (2006).
26. Ossowski, S., Schwab, R. & Weigel, D. Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J.* **53**, 674–690 (2008).
27. Lu, S. X., Knowles, S. M., Andronis, C., Ong, M. S. & Tobin, E. M. CIRCADIAN CLOCK ASSOCIATED1 and LATE ELONGATED HYPOCOTYL function synergistically in the circadian clock of *Arabidopsis*. *Plant Physiol.* **150**, 834–843 (2009).
28. Kikis, E. A., Khanna, R. & Quail, P. H. *ELF4* is a phytochrome-regulated component of a negative-feedback loop involving the central oscillator components CCA1 and LHY. *Plant J.* **44**, 300–313 (2005).
29. Sawa, M., Nusinow, D. A., Kay, S. A. & Imaizumi, T. FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis*. *Science* **318**, 261–265 (2007).
30. Para, A. *et al.* PRR3 is a vascular regulator of TOC1 stability in the *Arabidopsis* circadian clock. *Plant Cell* **19**, 3462–3473 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** P. Y. Pongsawakul illustrated the model in Fig. 4c. We thank G. Breton, T. Hirota, J. Pruneda-Paz, D. Nagel, E. Kolmos and B. J. Cole for critical reading of the manuscript. We also thank J. Halverson, A. L. Quiroz and C. Valdivia for excellent technical assistance. *pif4-101 pif5-1* seedlings were a generous gift from S. Lorrain and C. Fankhauser. F. Harmon originally identified the nature of the *elf4-3* mutation and designed the dCAPS strategy. This work was supported by a University of California, San Diego, Chancellor's Undergraduate Research Scholarship (to J.J.K.), grants from the European Molecular Biology Organization (ALTF 236-2005 to A.H.), the National Science Foundation (IBN-0416762 to T.F.S.) and the National Institutes of Health (NRSA GM083585 to D.A.N., NRSA GM080930 to E.E.H., R01 GM79712 to T.I., and R01 GM50006 and GM67837 to S.A.K.).

**Author Contributions** D.A.N., A.H., E.E.H., T.I., T.F.S., E.M.F. and S.A.K. designed the experiments. D.A.N. performed and analysed all of the immunoprecipitations and ChIP assays, generated and characterized the anti-*ELF3* antibody and the transgenic lines, and generated the plasmids for yeast three-hybrid analysis. D.A.N. also co-performed the gene expression analysis with A.H. A.H. performed the yeast two- and one-hybrid assays and generated and characterized the LUX and LUX/NOX ami transgenic lines. E.E.H. generated the anti-LUX antibody, characterized transgenic lines and co-performed western blot analysis with D.A.N. J.J.K. measured hypocotyls, performed and analysed the yeast three-hybrid assay, and assisted with the yeast one-hybrid analysis and generation of the *elf3-2 pif4-101 pif5-1* seedlings. T.I. performed the original yeast two-hybrid assays. T.F.S. generated the *ELF4::ELF4-HA elf4-2* line. E.M.F. characterized the *ELF4*-transgenic lines. D.A.N. and S.A.K. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.A.K. ([skay@ucsd.edu](mailto:skay@ucsd.edu)).



## METHODS

**Yeast one-hybrid analysis.** All reporter strains were generated by homologous recombination of pGLacZi constructs (Clontech) in the yeast strain YM4271, according to the manufacturer's instructions. pGLacZi is a Gateway-compatible version of pLacZi (Clontech)<sup>20</sup>. Promoter fragments were amplified using primers listed in Supplementary Table 1 and were cloned into pENTR/D-TOPO (Invitrogen) and then transferred to pGLacZi, according to the manufacturer's instructions. To generate translational fusions to GAL4-AD, the coding sequence of *LUX* was cloned into pENTR/D-TOPO and subsequently recombined into pACTGW as previously described<sup>20</sup>. Transformations of AD constructs into the reporter strains and determinations of the  $\beta$ -galactosidase ( $\beta$ -gal) activity were performed in a 96-well format as previously described<sup>20</sup>.  $\beta$ -Gal activities were normalized to the control with an empty pACTGW vector.

**Yeast two-hybrid analysis.** cDNAs encoding full-length *LUX* (described above), *ELF3*, *ELF4*, *LUX-N* (amino acids 1–143) and *LUX-C* (amino acids 144–324) were cloned into the pENTR/D-TOPO vector (Invitrogen) (Supplementary Table 1). After the sequences had been verified, they were transferred into the pACTGW vector by Gateway LR recombination reaction (Invitrogen) to generate the bait plasmids<sup>20</sup>. *ELF4* and *ELF3* cDNAs were transferred into pASGW by a Gateway LR recombination reaction (Invitrogen) to generate the prey plasmids. The detailed yeast two-hybrid procedure was as previously described<sup>29</sup>.

**Yeast three-hybrid analysis.** Yeast three-hybrid analysis was performed as described previously<sup>30</sup>, with the following modifications: *ELF3* with an amino-terminal FLAG-epitope tag was cloned from cDNA into a pENTR/D-TOPO vector using the primers 5'-CGCGGCCGCAAAATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGACGATGACAAAATGAAGAGAGGGAAAGATGAGGAG-3' and 5'-TTGGTTCTGCCATGAGACTG-3', and then inserted into the original pENTR/D-TOPO-*ELF3* clone using the restriction enzymes NotI and EcoRI (New England BioLabs) and confirmed by sequencing. *ELF4* was then cloned into the pBridge vector by amplifying with the primers 5'-GGGGGAATTCATGAAGAGGAACGGCGAGAC-3' and 5'-TTTCTGTCAGTTAAGCTCTAGTTCGGCAGC-3', and inserting into EcoRI and PstI (New England BioLabs) restriction sites. *FLAG-ELF3* was then cloned into either the pBridge vector or pBridge-*ELF4* using the restriction sites of NotI and EcoRV (New England BioLabs), after first digesting either the pBridge or pBridge-*ELF4* vector with BglI, blunting with Klenow and then digesting with NotI (New England BioLabs). pBridge-*ELF3* or pBridge-*ELF4-ELF3* was introduced into yeast strain YM4271 and then mated to strains containing the vector pACTGW, pACTGW-*LUX* or pACTGW-*NOX*<sup>20</sup> in the yeast strain AH109, according to the manufacturer's protocol (Clontech). Yeast were grown under selection and analysed for  $\beta$ -gal activity, as described by the manufacturer's instructions (Clontech) with the modifications for 96-well analysis<sup>20</sup>.

**Plant materials and growth conditions.** All wild-type, mutant and transgenic lines were in *A. thaliana* ecotype Columbia-0. *CAB2::LUC*-reporter-containing lines have been described previously<sup>31</sup>. Seeds were chlorine-gas sterilized and plated onto 1× Murashige and Skoog (MS) basal salt medium with 1.5% agar and 3% (w/v) sucrose. After stratification in the dark at 4 °C for 3 days, plates were transferred to an incubator (Percival Scientific) that was set to the indicated light conditions and a constant temperature of 22 °C. Light entrainment was in 12L:12D cycles or in short-day and long-day photoperiods (8L:16D and 16L:8D, respectively), with light supplied at 80  $\mu\text{mol m}^{-2} \text{s}^{-1}$  by cool-white fluorescent bulbs. To analyse seedling morphology, evenly spaced seedlings were grown under 12L:12D conditions at 22 °C and measured on day 10. Photographs of seedlings were analysed using NIH ImageJ software (<http://rsbweb.nih.gov/ij/>).

**Construction of double and triple mutants.** *ELF4::ELF4-HA elf3-1 elf4-2 CAB2::LUC* double mutants were generated by genetic crosses between *elf3-1* (ref. 14) and *elf4-2 ELF4::ELF4-HA #1* (Basta resistance) *CAB2::LUC*, and *F<sub>2</sub>* populations were screened for long hypocotyls, Basta resistance, luminescence and an arrhythmic bioluminescence phenotype in constant light. *elf4-2 (arr44)*<sup>21</sup> mutations were identified by dCAPS PCR method<sup>32</sup> using the primers 5'-ATGGGTTTGCTCCACGGATTA-3' and 5'-CAGGTTCCGGGAACCAAATTCT-3', and the restriction enzyme HpyCH4V (New England BioLabs) to analyse for the presence of the mutation. The *elf3-1* mutation was confirmed by 100% long hypocotyl, as well as by analysis using dCAPS primers 5'-TTGTCAGAGGATAAGCTGCGCT-3', 5'-TGTTGGCTGTTGCTGTTGCTGT-3' and the restriction enzyme HincII, and by loss of the *ELF3* signal in western blotting. *lux-4 elf3-1 CAB2::LUC* double mutants were generated by crossing *elf3-1* to *lux-4 CAB2::LUC*, and *F<sub>2</sub>* populations were screened for long hypocotyls, luminescence and an arrhythmic bioluminescence phenotype in constant light. Loss of *LUX* and *ELF3* was confirmed by assessing hypocotyl length, performing dCAPS PCR for *elf3-1* and *lux-4* (using the primers 5'-ATGGAGATGACGGTGGCGGT-3' and 5'-AACGAATCTCTTGTTAGCTGCGGAGT-3' and the restriction enzyme HinfI (New England BioLabs)), and carrying out

western blot analysis. *elf4-3 elf3-1 CAB2::LUC* double mutants were generated by crossing *elf4-3 CAB2::LUC*, which was generated by EMS mutagenesis as previously described<sup>11</sup>, with *elf3-1*, and these mutants were screened as above. The mutant *elf4-3* contains a single point mutation in the coding sequence of *ELF4* that results in a truncated protein (W26\*), which was identified by sequencing. The dCAPS primers 5'-GAGCAGGGAGAGGATCCAGCGATGTG-3' and 5'-CCGACGAGAACTAGTATTGA-3' and the restriction enzyme BstXI (New England BioLabs) were used to screen for the mutation. The presence of the *elf3-1* mutation was confirmed by dCAPS and western blotting. The *elf3-2* lines<sup>14</sup> were crossed to *TOC1::LUC* lines as described previously<sup>33</sup>, and we analysed *F<sub>2</sub>* populations for long hypocotyls and bioluminescence. The *elf3-2* mutation was mapped using the TAIL PCR method, which identified an inversion<sup>34</sup>. A PCR strategy over the inversion was used to distinguish wild-type lines from mutant lines using the following primers: 5'-TGAGTATTTGTTTCTTCGAGC-3' and 5'-CATATGGAGGGAAGTAGCCATTAC-3' for wild type and 5'-TGGTTATTATTCTCCGCTCTTTC-3' and 5'-TTGTTCCATTAGCTGTTCAACCTA-3' for *elf3-2*. The combination mutants *elf3-2 pif4-101 pif5-1 TOC1::LUC*, *elf3-2 pif5-1 TOC1::LUC*, *elf3-2 pif4-101 TOC1::LUC*, *pif4-101 pif5-1 TOC1::LUC*, *pif5-1 TOC1::LUC*, and *pif4-10 TOC1::LUC* were generated from crosses between *elf3-2 TOC1::LUC* and *pif4-101 pif5-1* double mutants. *F<sub>2</sub>* plants were screened for bioluminescence and then analysed for mutant backgrounds by PCR as previously described<sup>13</sup>. Homozygous *F<sub>3</sub>* populations were identified by screening for mutations and transgenes. The generation and characterization of *LUX::LUX-GFP lux-4 CAB2::LUC* has been described previously<sup>20</sup>.

**GFP and LUX/NOX ami line generation.** The coding sequence of GFP was amplified by PCR from the pK7FWG2 vector<sup>35</sup> using the following primers 5'-CACCATGTGGTCTCATCCTCAATTTGAAAAAGCGCGGTTGGTCTC ATCCTCAATTTGAAAAAGGTGGTATGGTGAGCAAGGCGGAGGAGCTG-3' and 5'-TCAAGCGTAATCTGGAACATCGTATGGGTACACATCCTGTAC AGCTCGTCCATGCC-3', which introduce a StrepII epitope (SII) tag to the N terminus and an HA tag to the carboxy terminus. This fragment was then cloned into Gateway pENTR/D-TOPO. After sequencing, this construct was recombined with the pB7WG2 vector<sup>35</sup> to constitutively express SII-GFP-HA under the control of the 35S promoter. This construct was introduced into *CCA1::LUC* lines<sup>36</sup> by *Agrobacterium*-mediated transformation<sup>37</sup>. Transformants were selected based on Basta resistance and fluorescence and were screened for single insertion. Lines were brought to homozygosity before use.

The amiRNA (TATGATATCCTTGGCTACCCA) targeting *LUX* and *NOX* was constructed as described previously<sup>25</sup>. Primers designed using WMD3 Web microRNA Designer (<http://wmd3.weigelworld.org/cgi-bin/webapp.cgi>) were used to amplify the amiRNA precursor by overlapping PCR from the pRS300 template. The fragment containing the amiRNA foldback was cloned into pENTR/D-TOPO, sequenced and subsequently recombined using Gateway LR Clonase II (Invitrogen) into the pB2GW7 vector<sup>35</sup> for constitutive expression under control of the 35S promoter. This construct was transformed into a *CAB2::LUC* reporter background<sup>31</sup> using *Agrobacterium* infiltration<sup>37</sup>. Transformants were selected on Basta, and all experiments were performed in single-insertion, homozygous plants.

**Luciferase imaging.** After 6 days of entrainment, plants were sprayed with 5 mM luciferin (Biosynth) prepared in 0.01% (v/v) Triton X-100 (Sigma-Aldrich) and transferred to constant light (80  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ) 1 day before imaging. The emitted luminescence was recorded every 2.5 h over 5 days, using a digital CCD camera (Hamamatsu Photonics). The images were processed using MetaMorph imaging software (Molecular Devices), and the data were analysed by fast Fourier transform–nonlinear least squares (FFT–NLLS)<sup>38</sup> using the interface provided by the Biological Rhythms Analysis Software System version 3.0 (BRASS) (<http://www.amiar.org>).

**Generation of anti-ELF3 antibody.** Antibodies were generated in rabbits (Sigma Genosys) against an ELF3-specific peptide, containing an additional N-terminal cysteine for conjugation (CSIQEERKRYDSSKP), corresponding to amino acids 681–694 of ELF3. Antibodies were affinity purified against this peptide using a SulfoLink Immobilization Kit (Thermo Scientific). Eluted antibody-containing fractions were buffer exchanged into 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 50% glycerol and 0.02% NaN<sub>3</sub> by using an equilibrated PD-10 column (GE Healthcare) and then stored at –80 °C.

**Generation of anti-LUX antibody.** Full-length *LUX* protein was expressed as a glutathione S-transferase (GST) fusion, which was purified and used to immunize rabbits to obtain polyclonal antisera (Open Biosystems). Antibodies were purified using an affinity column made of purified GST–*LUX* bound to Affi-Gel 15 Activated Immunoaffinity Support (Bio-Rad)<sup>39</sup>. Antibodies were eluted from the affinity column with 100 mM glycine, pH 2.5, exchanged into storage buffer (1× PBS, 50% glycerol and 0.02% NaN<sub>3</sub>) using a PD-10 buffer exchange column and then stored at –80 °C.

**Construction of *ELF4::ELF4-HA*.** *ELF4* was cloned from genomic DNA to include 580 bp of promoter sequence, using the primers 5'-CACCGTCTTGC ATAACATGAAGC-3' and 5'-AGCTCTAGTTCGGCAGCACC-3', and then cloned into Gateway pENTR/D-TOPO. After sequencing, this construct was recombined with pEarleyGate 301 to introduce a C-terminal HA tag to *ELF4* (ref. 40). This construct was introduced into *elf4-2 CAB2::LUC* lines by *Agrobacterium*-mediated transformation<sup>37</sup>. Transformants were selected based on Basta resistance and screened for single insertion. Lines were brought to homozygosity before use.

**ChIPs.** Roughly 5 g (fresh weight) whole seedlings were harvested and crosslinked for 10 min under vacuum in crosslinking buffer (10 mM Tris, pH 8.0, 1 mM EDTA, 250 mM sucrose, 1 mM PMSF and 1% formaldehyde). Crosslinking was quenched in 125 mM glycine, pH 8.0, under vacuum for 5 min, and then seedlings were washed three times in double-distilled water and rapidly frozen before disruption in a ball mill (Retsch) under liquid nitrogen. Ground tissue was processed as described previously<sup>41</sup>, with the following modifications: sucrose-gradient-purified nuclei were resuspended in SII buffer (100 mM Na-phosphate, pH 8.0, 150 mM NaCl, 5 mM EDTA, 5 mM EGTA, 0.1% Triton X-100, 1 mM PMSF and 1× protease inhibitor cocktail (Roche)) and sonicated (Branson) at 15% power, with 0.5 s on/off cycles for a total of 30 s on ice until the average chromatin size was ~ 500 bp. The extracts were clarified by centrifugation at 20,000g and stored at -80 °C until use. Technical replicates containing approximately 1.5 mg DNA were resuspended in 800 µl SII buffer, incubated with 2 µg anti-GFP antibody (ab290, Abcam), anti-HA antibody (3F10, Roche) or anti-ELF3 antibody bound to Protein G Dynabeads (Invitrogen) for 1.5 h at 4 °C and then washed five times with SII buffer. Chromatin was eluted from the beads twice at 65 °C with Stop buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 20 mM EDTA and 1% SDS). RNase- and DNase-free glycogen (2 µg) (Boehringer Mannheim) was added to the input and eluted chromatin before they were incubated with DNase- and RNase-free proteinase K (Invitrogen) at 65 °C overnight and then treated with 2 µg RNase A (Qiagen) for 1 h at 37 °C. DNA was purified by phenol-chloroform extraction, followed by two serial ethanol precipitations. Quantitative PCR reactions of the technical replicates were performed using the CFX384 Real Time PCR Detection System (Bio-Rad), with the following PCR conditions: 3 min at 95 °C, followed by 40 cycles of 10 s at 95 °C, 10 s at 55 °C and 20 s at 72 °C in a buffer consisting of 1× Ex Taq buffer (TaKaRa Bio), 0.5× SYBR Green (Molecular Probes), 5 nM fluorescein (Bio-Rad), 0.05% (v/v) Tween 20, 2.5% (v/v) DMSO, 25 µg ml<sup>-1</sup> BSA (New England BioLabs), 0.25 mM dNTPs, 250 nM primers and 1 U Taq DNA polymerase (BioPioneer). Primers used in this study are listed in Supplementary Table 1.

**Immunoprecipitations and western blots.** Approximately 500 mg whole seedlings were transferred to 2-ml tubes with three 3.2-mm stainless steel beads, and then frozen and disrupted in a ball mill under liquid nitrogen. After removing ~100 mg tissue for RNA analysis, ground tissue was resuspended in 400 µl SII buffer containing 1× phosphatase inhibitor cocktails 1 and 2 (Sigma) and 10 µM MG-132 (Peptides International) and sonicated twice at 10% power, with 0.5 s on/off cycles for a total of 20 s on ice. Extracts were then clarified by centrifugation at 4 °C, measured for protein concentration using 1× Bradford reagent (Bio-Rad) and normalized to 3 mg ml<sup>-1</sup> in SII buffer for western blots. For immunoprecipitations, extracts were diluted to 1.5 mg ml<sup>-1</sup> in SI buffer. Anti-HA antibody (4 µg) crosslinked to Protein G Dynabeads was added to extracts, rotated for 1.5 h at 4 °C, then washed 3× with SII buffer. Precipitated protein was eluted by heating beads at 65 °C for 5 min in 25 µl SDS-PAGE loading buffer. Protein levels were then analysed by western blot using either horse radish peroxidase (HRP)-conjugated 3F10 anti-HA (1:2,000, Roche), anti-ELF3 (1:750) or anti-LUX (1:750) antibody, followed by HRP-conjugated anti-rabbit secondary antibody (1:2,000, Pierce). The ACTIN loading control was detected using an anti-ACTIN mouse antibody, mAB1501 (1:2,000, Millipore), followed by alkaline-phosphatase-conjugated

anti-mouse secondary antibody (1:4,000, Promega). Blots for *ELF4* represent 20% of the total immunoprecipitation sample, because *ELF4* must be run on a separate, 15%, gel owing to its low molecular weight; these gels are noted by (\*). The dot (•) denotes a background signal arising from the crosslinked HA beads (data not shown). *LUX* runs as high- and low-molecular-weight isoforms, denoted by (-). When samples were collected in the dark, extracts were made and immunoprecipitations were assembled under a safe green light and protected from light until they were eluted in SDS-PAGE loading buffer before loading onto gels.

**Antibody crosslinking.** Antibody (2 µg) was crosslinked to 12 µl Protein G Dynabeads according to the manufacturer's instructions, with the following modifications: quenching of the dimethyl pimelimidate was performed with 0.2 M ethanolamine (pH 8.0), followed by two washes with 0.1 M glycine (pH 2.5) and neutralization with neutralization buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl and 0.01% Triton X-100), and the samples were then stored at -20 °C in storage buffer (50% glycerol, 50 mM Tris, pH 8.0, 150 mM NaCl, 0.01% Triton X-100 and 0.03% NaN<sub>3</sub>) until use.

**RNA extractions.** Seedlings were grown on Whatman filter paper atop MS plates under 12L:12D, 8L:16D or 16L:8D conditions and harvested on day 12, or were transferred to constant light on day 10 and harvested 2 to 3 days later. Total RNA was isolated using an RNeasy Plant Mini Kit (Qiagen). For cDNA synthesis, 1 µg total RNA was reverse-transcribed using the iScript cDNA synthesis kit (Bio-Rad). Synthesized cDNA was quantified by real-time quantitative PCR using the CFX-384 Real Time System (Bio-Rad), with the following PCR conditions: 3 min at 95 °C, followed by 40 cycles of 10 s at 95 °C, 10 s at 55 °C and 20 s at 72 °C in a buffer consisting of 1× ExTaq buffer, 1× SYBR Green, 10 nM fluorescein (Bio-Rad), 0.1% (v/v) Tween 20, 5% (v/v) DMSO, 50 µg ml<sup>-1</sup> BSA, 0.25 mM dNTPs, 250 nM primers and 1 U Taq DNA polymerase. Isopentenyl pyrophosphate/dimethylallyl pyrophosphate isomerase (*IPP2*) (At3g02780), ascorbate peroxidase 3 (*APX3*) (At4g35000) and aspartyl protease family protein (At1g1910) were used as the normalization controls<sup>11,19</sup>. Primer sequences are shown in Supplementary Table 1 and were designed using Primer3 (ref. 42) or as described for *PIF4* and *PIF5* (ref. 43).

31. Millar, A. J., Carre, I. A., Strayer, C. A., Chua, N. H. & Kay, S. A. Circadian clock mutants in *Arabidopsis* identified by luciferase imaging. *Science* **267**, 1161–1163 (1995).
32. Neff, M. M., Turk, E. & Kalishman, M. Web-based primer design for single nucleotide polymorphism analysis. *Trends Genet.* **18**, 613–615 (2002).
33. Alabadi, D. et al. Reciprocal regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* circadian clock. *Science* **293**, 880–883 (2001).
34. Liu, Y. G., Mitsukawa, N., Oosumi, T. & Whittier, R. F. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* **8**, 457–463 (1995).
35. Karimi, M., Inzé, D. & Depicker, A. GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* **7**, 193–195 (2002).
36. Pruneda-Paz, J., Breton, G., Para, A. & Kay, S. A. A functional genomics approach reveals CHE as a component of the *Arabidopsis* circadian clock. *Science* **323**, 1481–1485 (2009).
37. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
38. Plautz, J. D. et al. Quantitative analysis of *Drosophila* period gene transcription in living animals. *J. Biol. Rhythms* **12**, 204–217 (1997).
39. Harlow, E. & Lane, D. *Using Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratories Press, 1999).
40. Earley, K. W. et al. Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616–629 (2006).
41. Haring, M. et al. Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. *Plant Methods* **3**, 11 (2007).
42. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
43. Czechowski, T., Bari, R. P., Stitt, M., Scheible, W. R. & Udvardi, M. K. Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. *Plant J.* **38**, 366–379 (2004).

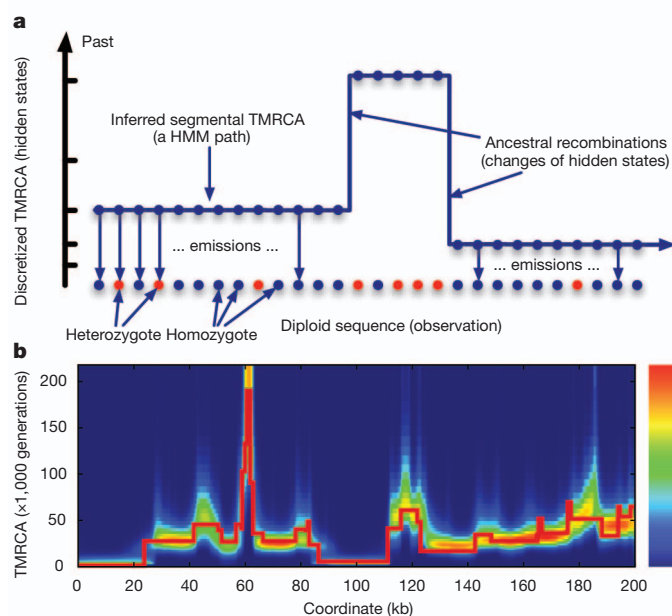
# Inference of human population history from individual whole-genome sequences

Heng Li<sup>1,2</sup> & Richard Durbin<sup>1</sup>

The history of human population size is important for understanding human evolution. Various studies<sup>1–5</sup> have found evidence for a founder event (bottleneck) in East Asian and European populations, associated with the human dispersal out-of-Africa event around 60 thousand years (kyr) ago. However, these studies have had to assume simplified demographic models with few parameters, and they do not provide a precise date for the start and stop times of the bottleneck. Here, with fewer assumptions on population size changes, we present a more detailed history of human population sizes between approximately ten thousand and a million years ago, using the pairwise sequentially Markovian coalescent model applied to the complete diploid genome sequences of a Chinese male (YH)<sup>6</sup>, a Korean male (SJK)<sup>7</sup>, three European individuals (J. C. Venter<sup>8</sup>, NA12891 and NA12878 (ref. 9)) and two Yoruba males (NA18507 (ref. 10) and NA19239). We infer that European and Chinese populations had very similar population-size histories before 10–20 kyr ago. Both populations experienced a severe bottleneck 10–60 kyr ago, whereas African populations experienced a milder bottleneck from which they recovered earlier. All three populations have an elevated effective population size between 60 and 250 kyr ago, possibly due to population substructure<sup>11</sup>. We also infer that the differentiation of genetically modern humans may have started as early as 100–120 kyr ago<sup>12</sup>, but considerable genetic exchanges may still have occurred until 20–40 kyr ago.

The distribution of the time since the most recent common ancestor (TMRCAs) between two alleles in an individual provides information about the history of change in population size over time. Existing methods for reconstructing the detailed TMRCAs distribution have analysed large samples of individuals at non-recombining loci like mitochondrial DNA<sup>13</sup>. However, the statistical resolution of inferences from any one locus is poor, and power fades rapidly upon moving back in time because there are few independent lineages probing deep time depths (in humans, no information is available from mitochondrial DNA beyond about 200 kyr ago, when all humans share a common maternal ancestor<sup>11</sup>). In contrast, a diploid genome sequence contains hundreds of thousands of independent loci, each with its own TMRCAs between the two alleles carried by an individual. In principle, it should be possible to reconstruct the TMRCAs distribution across the autosomes and the X chromosome by studying how the local density of heterozygous sites changes across the genome, reflecting segments of constant TMRCAs separated by historical recombination events. To explore whether we could use this idea to learn about the detailed TMRCAs distribution from a diploid whole-genome sequence, we proposed the pairwise sequentially Markovian coalescent (PSMC) model, which is a specialization to the case of two chromosomes of the sequentially Markovian coalescent model<sup>14</sup> (Fig. 1a). The free parameters of this model include the scaled mutation rate, the recombination rate and piecewise constant ancestral population sizes (see Methods). We scaled results to real time, assuming 25 years per generation and a neutral mutation rate of  $2.5 \times 10^{-8}$  per generation<sup>15</sup>. The consequences of uncertainty in the two scaling parameters will be discussed later in the text.

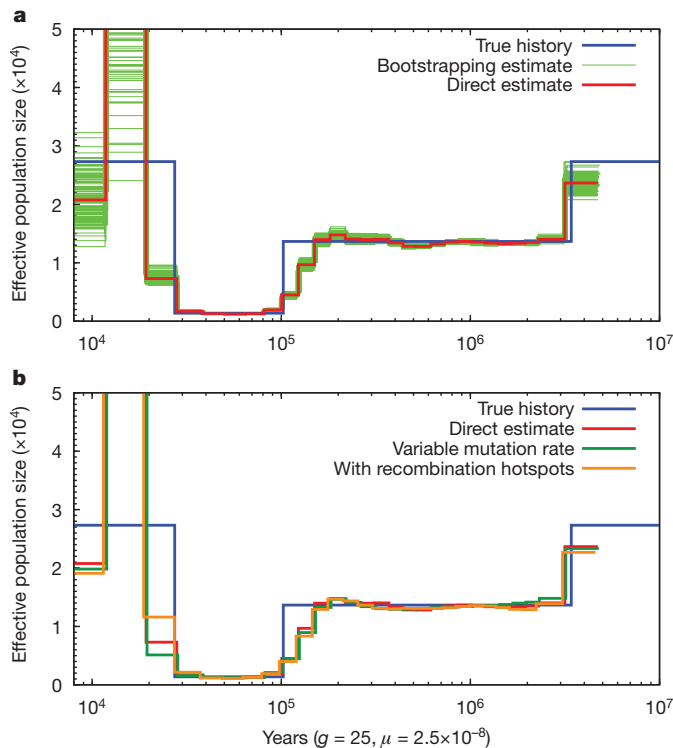
To validate our model, we simulated one-hundred 30-megabase (Mb) sequences with a sharp out-of-Africa bottleneck followed by a population expansion, and inferred population-size history with PSMC (Fig. 2a). PSMC was able to recover the parameters used in the simulation and the variance of the estimate was small between 20 kyr ago and 3 Myr ago. More recently than 20 kyr ago or more anciently than 3 Myr ago, few recombination events are left in the present sequence, which reduces the power of PSMC. Therefore, the estimated effective population size ( $N_e$ ) in these time intervals was not as accurate and had large variance. To test the robustness of the model, we introduced variable mutation rates and recombination hotspots in the simulation (Supplementary Information). The inference was still close to the true history (Fig. 2b) and a uniform rate of single nucleotide polymorphism (SNP) ascertainment errors did not change our qualitative results either (Supplementary Fig. 2). The simulations did, however, reveal a limitation of PSMC in recovering sudden changes in effective population size. For example, the instantaneous reduction from 12,000 to 1,200 at 100 kyr ago in the simulation was spread over several preceding tens of thousands of years in the PSMC reconstruction.



**Figure 1 | Illustration of the PSMC model and its application to simulated data.** **a**, The PSMC infers the local time to the most recent common ancestor (TMRCAs) on the basis of the local density of heterozygotes, using a hidden Markov model in which the observation is a diploid sequence, the hidden states are discretized TMRCAs and the transitions represent ancestral recombination events. **b**, We used the ms software to simulate the TMRCAs relating the two alleles of an individual across a 200-kb region (the thick red line), and inferred the local TMRCAs at each locus using the PSMC (the heat map). The inference usually includes the correct time, with the greatest errors at transition points.

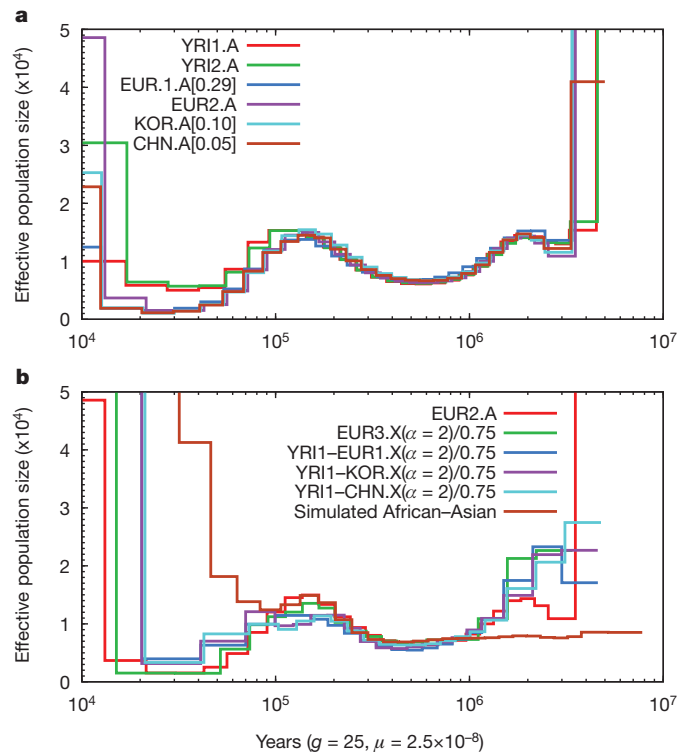
<sup>1</sup>The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. <sup>2</sup>Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA.





**Figure 2 | PSMC estimate on simulated data.** **a**, PSMC estimate on data simulated by msHOT. The blue curve is the population-size history used in simulation; the red curve is the PSMC estimate on the originally simulated sequence; the 100 thin green curves are the PSMC estimates on 100 sequences randomly resampled from the original sequence. **b**, PSMC estimate on data with a variable mutation rate or with hotspots.  $g$ , generation time;  $\mu$ , mutation rate.

We applied the PSMC model to real data from recently published genome sequences (see Table 1, which defines the acronyms for samples used elsewhere in the text and figures). Figure 3a shows that all populations are very similar in their estimated  $N_e$  history between 150 and 1,500 kyr ago. The Yoruba (YRI) genome differentiates from non-African populations around 100–120 kyr ago (at 110 kyr ago,  $N_e^{\text{YRI}} = 15,313 \pm 559$  and  $N_e^{\text{CHN}} = 12,829 \pm 485$ ). This evidence of early population differentiation is potentially consistent with the archaeological evidence of anatomically modern humans found in the Near East around 100 kyr ago<sup>12</sup>. European and East Asian populations are nearly identical in estimated  $N_e$  before 11 kyr ago. From a peak of 13,500 at 150 kyr ago, the  $N_e$  dropped by a factor of ten to 1,200 between 40 and 20 kyr ago, before a sharp increase, the precise magnitude of which we do not have the power to measure. We also observed a less marked bottleneck in YRI from a peak of 16,100 around 100–150 kyr ago to 5,700 at 50 kyr ago, recovering earlier<sup>16</sup> than the out-of-Africa populations, with an increase back to 8,700 by 20 kyr



**Figure 3 | PSMC estimate on real data.** **a**, Population sizes inferred from autosomes of six individuals. 5%, 10% and 29% of heterozygotes are assumed to be missing in CHN.A, KOR.A and EUR1.A, respectively. **b**, Population sizes inferred from male-combined X chromosomes and the simulated African-Asian combined sequences from the best-fit model in ref. 21. Sizes inferred from X-chromosome data are scaled by 4/3. The neutral mutation rate on X, which is used in time-scaling, is estimated with the ratio of male-to-female mutation rate,  $\alpha$ , equal to 2 (see Methods).

ago, coinciding with the Last Glacial Maximum. All populations showed increased  $N_e$  between 60 and 200 kyr ago, about the time of origin of anatomically modern humans<sup>17</sup>. An alternative to an increase in actual population size during this time would be that there was population structure involving separation and admixture<sup>11,16</sup> (Supplementary Fig 5).

We also saw an increase in estimated  $N_e$  before 1 million years (Myr) ago in all populations, with a sharp increase before 3 Myr ago. Although it is tempting to read into this the transition from the previously estimated larger  $N_e$  at the time of the split from the chimpanzee<sup>18</sup>, our method may also be subject to artefacts in this region, due to regions of balancing selection or to clustered false heterozygotes related to segmental duplications (Supplementary Fig. 3).

Analysis of a European female X chromosome (EUR3.X) yielded a history similar to that from autosomes scaled by 0.75, as expected for the X chromosome (Fig. 3b). We did not observe a more severe

**Table 1 | Properties of the input sequences**

Label	Description	Coverage	Number of called bases (bp)	Number of heterozygotes (bp)	Heterozygosity ( $\times 1,000$ )
YRI1.A (ref. 10)	NA18507 autosomes	$\times 40$	$2.14 \times 10^9$	$2.17 \times 10^6$	1.013
YRI2.A (ref. 9)	NA19239 autosomes	$\times 29$	$2.11 \times 10^9$	$2.21 \times 10^6$	1.051
EUR1.A (ref. 8)	Venter autosomes	$\times 9$	$2.13 \times 10^9$	$1.23 \times 10^6$	0.578
EUR2.A (ref. 9)	NA12891 autosomes	$\times 38$	$2.11 \times 10^9$	$1.67 \times 10^6$	0.791
KOR.A (ref. 7)	SJK autosomes	$\times 20$	$2.13 \times 10^9$	$1.47 \times 10^6$	0.690
CHN.A (ref. 6)	YH autosomes	$\times 30$	$2.19 \times 10^9$	$1.52 \times 10^6$	0.694
YRI3.X (ref. 9)	NA19240 X chromosome	$\times 38$	$1.06 \times 10^8$	$7.16 \times 10^4$	0.673
EUR3.X (ref. 9)	NA12878 X chromosome	$\times 35$	$1.10 \times 10^8$	$4.80 \times 10^4$	0.436
KOR-CHN.X	SJK-YH combined X chromosome	-	$1.02 \times 10^8$	$3.97 \times 10^4$	0.390
YRI1-EUR1.X	NA18507-Venter combined X chromosome	-	$0.83 \times 10^8$	$5.56 \times 10^4$	0.670
YRI1-KOR.X	NA18507-KOR combined X chromosome	-	$1.00 \times 10^8$	$6.69 \times 10^4$	0.669
YRI1-CHN.X	NA18507-YH combined X chromosome	-	$1.06 \times 10^8$	$6.95 \times 10^4$	0.657

Coverage equals the average number of reads covering HapMap3 loci. A base is said to be called if it passes all filters described (see Methods). The relatively lower coverage for EUR1.A leads to higher sampling bias at heterozygotes, which leads to underestimated heterozygosity, but this can be corrected by adjusting the neutral mutation rate in scaling (Supplementary Information, section 1.2).

bottleneck on the X chromosome<sup>19</sup>. To investigate the relationship between African and non-African populations, we combined X chromosomes from YRI and a non-African to construct a pseudo-diploid genome. From Fig. 3b, we can see that although African and non-African populations might have started to differentiate as early as 100–120 kyr ago, they largely remained as one population until approximately 60–80 kyr ago, the time point at which the YRI1–EUR1.X curve clearly leaves EUR3.X. This supports the recent analysis of the relationship between the Neanderthal genome and that of modern humans<sup>20</sup>, which concluded that West Africans and non-Africans descended from a homogeneous ancestral population in the last 100,000 years, with subsequent minor admixture out of Africa from Neanderthals, rather than an alternative explanation involving ancient (>300,000-year-old) sub-structure separating West African and non-African populations.

From Fig. 3b, it is also notable that there is a low  $N_e$  between African and non-African populations until approximately 20 kyr ago, indicating that there were substantial genetic exchanges between these populations long after the initial separation. Complete separation would correspond to very large or effectively infinite  $N_e$ , as seen more recently than 20 kyr ago. To explore whether the inferred recent gene flow is a modelling artefact, we simulated complete divergence at 60 kyr ago according to the model in ref. 21, and saw increased rather than reduced  $N_e$  in the period 20–60 kyr ago (brown line in Fig. 3b). To explore further, we extracted segments from YRI1–KOR.X that coalesced more recently than 50 kyr ago, according to PSMC. These comprised 220 segments covering 31.2 Mb (>20% of the X chromosome). We observed 1,363 base-pair (bp) differences in 20.7 Mb of call-able sequence in these segments, corresponding to an average divergence time of 37.4 kyr ago. In contrast, if we apply the same process to the simulated data from the model of ref. 21, the segments that PSMC identifies as having diverged more recently than 50 kyr ago cover only 0.4% of the simulated chromosome. The human–macaque divergence in the 220 segments was only 4% lower than the chromosome average, so regional variability in mutation rates cannot explain these results. In summary, the existence of long segments of low divergence between YRI1 and KOR supports the inference from PSMC that there was substantial genetic exchange between West African and non-African populations up until 20–40 kyr ago, and is not consistent with a simple separation approximately 60 kyr ago.

The time frame proposed above for continued genetic exchange between Africans and non-Africans is more recent than the archaeologically documented time of the out-of-Africa dispersal, because there are modern human fossils in both Europe and Australasia that date to >40 kyr ago<sup>22</sup>. Further analysis of additional non-African genomes indicates that this genetic exchange occurred primarily before the separation of Europeans and East Asians (Supplementary Information, section 4.3). An important caveat to this conclusion is the uncertainty of the per-year mutation rate of  $1.0 \times 10^{-9}$  ( $2.5 \times 10^{-8}/25$ ). Although this mutation rate agrees well with the rates estimated between primates averaged over millions of years (Supplementary Information, section 3.1), generation intervals as high as 29 years per generation over the last few thousand years<sup>23</sup>, and present mutation rates lower than  $2.5 \times 10^{-8}$  per generation<sup>9</sup>, are possible in principle. These factors could make our recent date estimates too recent, although it seems unlikely that such inaccuracies would be consistent with a date of final genetic exchange as far back as 60 kyr ago. Our analyses also cannot exclude the possibility that the divergence time inferred from X chromosomes may not be representative, owing to sex-biased demographic processes<sup>19</sup>, highlighting the importance of repeating this analysis on autosomal data once haploid whole-genome sequences become available<sup>24</sup>. Notably, a recent study using an orthogonal type of data (analysis of allele frequencies) also inferred that gene flow between Africans and non-Africans continued well after the initial out-of-Africa migration: in the case of that study, until 17–26 kyr ago<sup>25</sup>. An important goal for future work is to determine whether these recent dates reflect real

history, and if so, to obtain more detail about the timing and scale of the events involved.

In this paper we have introduced a method to infer the history of effective population size from genome-wide diploid sequence data. It is relatively straightforward to apply, with less potential ascertainment bias than existing methods that use selective genotyping data or resequencing data from a few loci. Furthermore, our method is computationally tractable and typically uses much more primary sequence data than the existing methods, which allows us to estimate population size at each time going back in history, rather than assuming a parametric structure of times, divergences and size changes. The results described above concerning the timing and depth of the out-of-Africa bottleneck are broadly consistent with previous studies, although our results are more detailed (Supplementary Information, section 4.2). The hypothesis that there was significant ongoing genetic exchange throughout the bottleneck is surprising in light of current views about human migrations; however, it is not inconsistent with the archaeological literature, and should motivate further research. There is the potential to extend this type of sequentially Markovian coalescent hidden Markov model approach to data from several individuals, which would access more recent times, but this will require inference over a substantially more complex hidden-state-space of trees on the haplotypes, with each Markov path representing an ancestral recombination graph<sup>14</sup>. In addition, there is the potential to apply the method to investigate the population-size history of other species for which a single diploid genome sequence has been obtained (Supplementary Information, section 2.2).

## METHODS SUMMARY

Illumina short reads were obtained from the NCBI Sequence Read Archive and capillary reads from TraceDB. Reads were aligned to the human reference genome with BWA<sup>26</sup>. The consensus sequences were called by SAMtools<sup>27</sup> and then divided into non-overlapping 100-bp bins, with a bin being scored as heterozygous if there is a heterozygote in the bin, or as homozygous otherwise. The resultant bin sequences were taken as the input of the PSMC estimate. Coalescent simulation was done by ms<sup>28</sup> and cosi<sup>21</sup> software. The simulated sequences were binned in the same way.

The free parameters in the discrete PSMC–HMM model are the scaled mutation rate, recombination rate and piecewise constant population sizes. The time interval spanned by each size parameter was manually chosen. The expectation-maximization iteration started from a constant-sized population history. The expectation step was done analytically; Powell's direction set method was used for the maximization step. Parameter values stabilized by the twentieth iteration and these were taken as the final estimate. All parameters were scaled to a constant that is further determined under the assumption of a neutral mutation rate,  $2.5 \times 10^{-8}$ .

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 1 April 2009; accepted 20 May 2011.**

**Published online 13 July 2011.**

1. Reich, D. E. *et al.* Linkage disequilibrium in the human genome. *Nature* **411**, 199–204 (2001).
2. Marth, G. T., Czabarka, E., Muvai, J. & Sherry, S. T. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics* **166**, 351–372 (2004).
3. Plagnol, V. & Wall, J. D. Possible ancestral structure in human populations. *PLoS Genet.* **2**, e105 (2006).
4. Keinan, A., Mullikin, J. C., Patterson, N. & Reich, D. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nature Genet.* **39**, 1251–1255 (2007).
5. Fagundes, N. J. R. *et al.* Statistical evaluation of alternative models of human evolution. *Proc. Natl Acad. Sci. USA* **104**, 17614–17619 (2007).
6. Wang, J. *et al.* The diploid genome sequence of an Asian individual. *Nature* **456**, 60–65 (2008).
7. Ahn, S.-M. *et al.* The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res.* **19**, 1622–1629 (2009).
8. Levy, S. *et al.* The diploid genome sequence of an individual human. *PLoS Biol.* **5**, e254 (2007).
9. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
10. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).

11. Behar, D. M. *et al.* The dawn of human matrilineal diversity. *Am. J. Hum. Genet.* **82**, 1130–1140 (2008).
12. Mellars, P. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* **313**, 796–800 (2006).
13. Atkinson, Q. D., Gray, R. D. & Drummond, A. J. mtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory. *Mol. Biol. Evol.* **25**, 468–474 (2008).
14. McVean, G. A. T. & Cardin, N. J. Approximating the coalescent with recombination. *Phil. Trans. R. Soc. B* **360**, 1387–1393 (2005).
15. Nachman, M. W. & Crowell, S. L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297–304 (2000).
16. Mellars, P. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proc. Natl Acad. Sci. USA* **103**, 9381–9386 (2006).
17. Wall, J. D. & Hammer, M. F. Archaic admixture in the human genome. *Curr. Opin. Genet. Dev.* **16**, 606–610 (2006).
18. Hobolth, A., Christensen, O. F., Mailund, T. & Schierup, M. H. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet.* **3**, e7 (2007).
19. Keinan, A., Mullikin, J. C., Patterson, N. & Reich, D. Accelerated genetic drift on chromosome X during the human dispersal out of Africa. *Nature Genet.* **41**, 66–70 (2009).
20. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
21. Schaffner, S. F. *et al.* Calibrating a coalescent simulation of human genome sequence variation. *Genome Res.* **15**, 1576–1583 (2005).
22. Mellars, P. A new radiocarbon revolution and the dispersal of modern humans in Eurasia. *Nature* **439**, 931–935 (2006).
23. Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–423 (2005).
24. Kitzman, J. O. *et al.* Haplotype-resolved genome sequencing of a Gujarati Indian individual. *Nature Biotechnol.* **29**, 59–63 (2010).
25. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).
26. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
27. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
28. Hudson, R. R. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to D. Bentley (Illumina) and J. Wang (Beijing Genomics Institute) for early access to the sequencing data. We thank A. Coghlan for the idea of bootstrapping, and N. Patterson, M. Przeworski, D. Reich, and members of the Durbin research group for discussions and critiques. This work was funded by Wellcome Trust grant WT077192.

**Author Contributions** R.D. proposed the basic strategy and designed the overall study. H.L. developed the theory, implemented the algorithm and analysed results. R.D. and H.L. wrote the manuscript.

**Author Information** The PSMC software package is freely available at <http://github.com/lh3/psmc>. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to R.D. ([rd@sanger.ac.uk](mailto:rd@sanger.ac.uk)) or H.L. ([lh3@sanger.ac.uk](mailto:lh3@sanger.ac.uk)).



## METHODS

**Read alignment.** Alignment for individuals from the 1000 Genomes Project (NA12878, NA12891, NA19239 and NA19240) was obtained from the project FTP site. Illumina sequence reads for NA18507, YH and SJK were obtained from the NCBI Sequence Read Archive (AC:ERA000005, SRA000271 and SRA008175, respectively) and Sanger sequencing reads for Craig Venter were obtained from NCBI TraceDB. These sequence reads were mapped by BWA<sup>26</sup> (0.5.5) against the human reference genome build 36, including unassembled contigs and the genome of Epstein–Barr virus (AC:NC\_007605), with pseudoautosomal regions on the Y chromosome masked. For Illumina short reads, BWA option ‘-q 15’ was applied to enable trimming of low-quality bases at the 3’ end. Base qualities of SJK reads were overestimated and were therefore recalibrated using GATK<sup>29</sup> after alignment, with known SNPs in dbSNP-129 discarded. For capillary reads, the BWA-SW algorithm with the default options was used.

**Calling the consensus sequence.** The diploid consensus sequence for an autosome was obtained by the ‘pileup’ command of the SAMtools software package<sup>27</sup>, and then processed with the following loci marked as missing data: 1) read depth is more than twice or less than half of the average read depth estimated on HapMap3 genotyping loci; 2) the root mean squared mapping quality of reads covering the locus is below 25; 3) the locus is within 10 bp around predicted short insertions or deletions; 4) the inferred consensus quality is below a threshold (20 for Illumina data and 10 for capillary data); 5) fewer than 18 out of the 35 overlapping 35-bp oligonucleotides from the reference sequence can be mapped elsewhere with zero or one mismatch.

The X-chromosome consensus was derived in a similar way but with pseudoautosomal regions filtered as missing data. The X chromosomes of males are haploid and therefore the few heterozygotes that were called were discarded as errors. The pseudo-diploid X chromosomes of males were combined by marking a difference as a heterozygote.

The consensus sequences were further divided into 100-bp non-overlapping bins with each bin represented as ‘missing’ (marked ‘.’) if  $\geq 90$  bases were filtered or uncalled; as heterozygous (‘1’) if  $> 10$  bp were called and there was at least one heterozygote; or as homozygous (‘0’) otherwise. The sequence of bin values was taken as the input of the PSMC inference.

**Coalescent simulation.** One-hundred sequences of 30 Mb were simulated by ms software<sup>28</sup> with piecewise constant history, as shown in Fig. 2a. To simulate variation in mutation rate, the local mutation rate averaged in a 20-kb window between human and macaque was calculated from the EPO cross-species alignment obtained from Ensembl v50. In the simulation, the local coalescent trees were simulated with ms but mutations were generated on the basis of the relative local mutation rate on a 30-Mb segment randomly drawn from the human–macaque alignment. The program msHOT was used to simulate sequences with recombination hotspots. The location and size of hotspots were randomly drawn from the hotspot map obtained from HapMap (release 21); the scaled recombination rate in hotspots was tenfold higher than that in non-hotspot regions.

The cosi software package was used to simulate sequences under the best-fit model from ref. 21. This model considers variable recombination rates, recombination hotspots and migration between African and non-African populations.

**Overview of the PSMC model.** In the PSMC–HMM, the observation is a binary sequence of ‘0’, ‘1’ and ‘.’, as described above. The emission probability from state  $t$  is  $e(1|t) = e^{-\theta t}$ ,  $e(0|t) = 1 - e^{-\theta t}$  and  $e(.|t) = 1$ ; the transition probability from  $s$  to  $t$  is:

$$p(t|s) = (1 - e^{-\rho t})q(t|s) + e^{-\rho s}\delta(t - s)$$

where  $\theta$  is the scaled mutation rate,  $\rho$  is the scaled recombination rate,  $\delta(\cdot)$  is the Dirac delta function and

$$q(t|s) = \frac{1}{\lambda(t)} \int_0^{\min\{s,t\}} \frac{1}{s} \times e^{-\int_u^t \frac{du}{\lambda(u)}} du$$

is the transition probability conditional on there being a recombination event, where  $\lambda(t) = N_e(t)/N_0$  is the relative population size at state  $t$ . The discrete-state HMM is constructed by dividing coalescence-time into intervals and integrating emission and transition probabilities in the intervals, which can be done analytically given a piecewise-constant function,  $\lambda(t)$ . The stationary distribution of TMRCA can also be analytically derived. Details are in Supplementary Information.

**Scaling to real time.** The estimated TMRCA is in units of  $2N_0$  time, and  $\lambda(t)$  is scaled to  $N_0$  as well. The value of  $N_0$  cannot be determined from the model itself. To estimate  $N_0$ , a neutral mutation rate  $\mu_A = 2.5 \times 10^{-8}$  on autosomes<sup>15</sup> was used and thus  $N_0^A = \theta/4\mu_A$ . Given the ratio of male-to-female mutation rate<sup>30</sup>  $\alpha = 2$ , the neutral mutation rate of X chromosomes was derived as  $\mu_X = \mu_A [2(2 + \alpha)]/[3(1 + \alpha)] = 2.2 \times 10^{-8}$ . If heterozygotes are missed uniformly at a probability  $p$ , this is equivalent to reducing the neutral mutation rate from  $\mu$  to  $\mu' = \mu(1 - p)$ . False negatives due to the lack of coverage can thus be corrected. Generations were converted to years under the assumption of 25 years per generation.

**Parameter estimate with PSMC.** Given a maximum TMRCA in the  $2N_0$  scale of  $T_{\max}$ , and a number of atomic time intervals  $n$ , let the boundaries of these intervals be  $t_i = 0.1 \exp[i/n \log(1 + 10T_{\max})] - 0.1$ ,  $i = 0, \dots, n$ . To reduce the complexity of the search space, blocks of adjacent atomic intervals were combined to have the same population-size parameter via a user-specified pattern. On autosome and simulated data,  $T_{\max} = 15$ ,  $n = 64$  and the pattern is ‘1\*4 + 25\*2 + 1\*4 + 1\*6’, which means that the first population-size parameter spans the first four atomic time intervals, each of the next 25 parameters spans two intervals, the twenty-seventh parameter spans four intervals and the last parameter spans the last six time intervals. On X-chromosome data,  $T_{\max} = 15$ ,  $n = 60$  and the pattern is ‘1\*6 + 2\*4 + 1\*3 + 13\*2 + 1\*3 + 2\*4 + 1\*6’.

In the expectation-maximisation (EM) parameter estimate, the initial population-size parameters were all set as 1, representing a constant-sized history, the scaled mutation rate was calculated to match the observed heterozygosity and the initial value of the scaled recombination rate was arbitrarily set as one-quarter of the mutation rate. At the maximization step, Powell’s direction set method was used to minimize the  $Q$  function in the EM algorithm numerically. Parameters at the twentieth EM iteration were taken as the final results.

Bootstrapping was applied by breaking the consensus sequences into 5-Mb segments and randomly sampling a set of segments with replacement, such that the total length of the sampled segments was close to the size of the human reference genome.

Further discussion of methods and parameters is given in Supplementary Information.

29. McKenna, A. *et al.* The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
30. Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K. & Yasunaga, T. Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 863–867 (1987).

# The crystal structure of a voltage-gated sodium channel

Jian Payandeh<sup>1</sup>, Todd Scheuer<sup>1</sup>, Ning Zheng<sup>1,2</sup> & William A. Catterall<sup>1</sup>

Voltage-gated sodium (Na<sub>v</sub>) channels initiate electrical signalling in excitable cells and are the molecular targets for drugs and disease mutations, but the structural basis for their voltage-dependent activation, ion selectivity and drug block is unknown. Here we report the crystal structure of a voltage-gated Na<sup>+</sup> channel from *Arco bacter butzleri* (NavAb) captured in a closed-pore conformation with four activated voltage sensors at 2.7 Å resolution. The arginine gating charges make multiple hydrophilic interactions within the voltage sensor, including unanticipated hydrogen bonds to the protein backbone. Comparisons to previous open-pore potassium channel structures indicate that the voltage-sensor domains and the S4–S5 linkers dilate the central pore by pivoting together around a hinge at the base of the pore module. The NavAb selectivity filter is short, ~4.6 Å wide, and water filled, with four acidic side chains surrounding the narrowest part of the ion conduction pathway. This unique structure presents a high-field-strength anionic coordination site, which confers Na<sup>+</sup> selectivity through partial dehydration via direct interaction with glutamate side chains. Fenestrations in the sides of the pore module are unexpectedly penetrated by fatty acyl chains that extend into the central cavity, and these portals are large enough for the entry of small, hydrophobic pore-blocking drugs. This structure provides the template for understanding electrical signalling in excitable cells and the actions of drugs used for pain, epilepsy and cardiac arrhythmia at the atomic level.

Electrical signals (termed action potentials) encode and process information within the nervous system and regulate a wide range of physiological processes<sup>1,2</sup>. The voltage-gated ion channels (VGICs) that mediate electrical signalling have distinct functional roles<sup>1,2</sup>. Na<sub>v</sub> channels initiate action potentials. Voltage-gated calcium (Ca<sub>v</sub>) channels initiate processes such as synaptic transmission, muscle contraction and hormone secretion in response to membrane depolarization. Voltage-gated potassium (K<sub>v</sub>) channels terminate action potentials and return the membrane potential to its resting value. Na<sub>v</sub> channels are mutated in inherited epilepsy, migraine, periodic paralysis, cardiac arrhythmia and chronic pain syndromes<sup>3</sup>. These channels are molecular targets of drugs used in local anaesthesia and in the treatment of genetic and sporadic Na<sub>v</sub> channelopathies in the brain, skeletal muscle and heart<sup>4</sup>. The rapid activation, Na<sup>+</sup> selectivity and drug sensitivity of Na<sub>v</sub> channels are unique among VGICs<sup>2</sup>.

VGICs share a conserved architecture in which four subunits or homologous domains create a central ion-conducting pore surrounded by four voltage sensors<sup>5</sup>. The voltage-sensing domain (VSD) is composed of the S1–S4 segments, and the pore module is formed by the S5 and S6 segments with a P-loop between them<sup>5</sup>. The S4 segments place charged amino acids within the membrane electric field that undergo outward displacement in response to depolarization and initiate opening of the central pore<sup>6,7</sup>. Although the architecture of K<sub>v</sub> channels has been established at high resolution<sup>8,9</sup>, the structural basis for rapid, voltage-dependent activation of VGICs remains uncertain<sup>7,9</sup>, and the structures responsible for Na<sup>+</sup>-selective conductance and drug block in Na<sub>v</sub> channels are unknown. The primary pore-forming subunits of Na<sub>v</sub> and Ca<sub>v</sub> proteins in vertebrates are composed of approximately 2,000 amino acid residues in four linked homologous domains<sup>5</sup>. The bacterial NaChBac channel family is an important model for structure–function studies of more complex vertebrate Na<sub>v</sub> and Ca<sub>v</sub> channels<sup>10,11</sup>. NaChBac is a homotetramer, and its pharmacological profile is similar to Na<sub>v</sub> and Ca<sub>v</sub>

channels<sup>10,12</sup>. Bacterial Na<sub>v</sub> channels are highly Na<sup>+</sup> selective, but they can be converted into Ca<sup>2+</sup>-selective forms through simple mutagenesis<sup>13</sup>. The NaChBac family represents the probable ancestor of vertebrate Na<sub>v</sub> and Ca<sub>v</sub> channels. Through analysis of the three-dimensional structure of NavAb from *A. butzleri*, we provide the first insights into the structural basis of voltage-dependent gating, ion selectivity and drug block in Na<sub>v</sub> and Ca<sub>v</sub> channels.

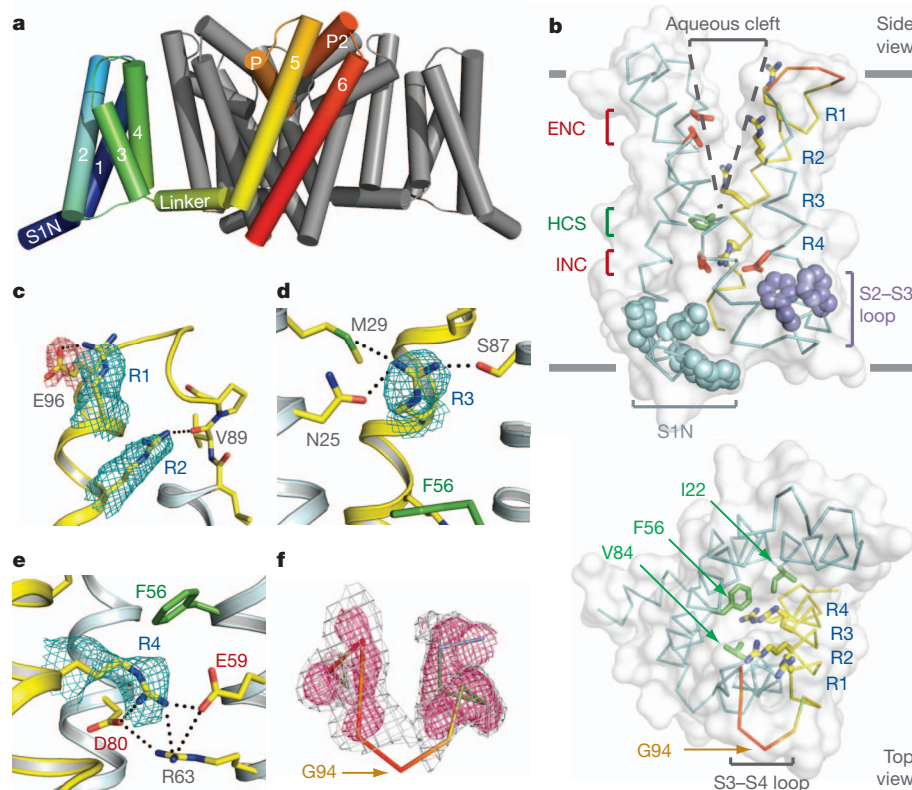
## Structure of NavAb in a membrane environment

NavAb is a member of the NaChBac family and functions as a voltage-gated sodium-selective ion channel (Supplementary Figs 1 and 2). Vertebrate Ca<sub>v</sub> channels require solubilization in digitonin and Na<sub>v</sub> channels require specific lipids to retain function when purified<sup>14,15</sup>. Accordingly, we solubilized NavAb in digitonin, crystallized it in a lipid-based bicelle system, and determined its structure at 2.70 Å resolution (Supplementary Figs 3–6 and Supplementary Table 1). NavAb crystallized as a dimer-of-dimers with 28 lipid molecules bound per tetramer (Supplementary Figs 3 and 6b). Crystal packing indicates a membrane-like environment (Supplementary Fig. 6a). NavAb VSDs interact noncovalently with the pore module of a neighbouring subunit (Fig. 1a), and crystallographic temperature factors highlight their dynamic nature (Supplementary Fig. 6c).

## Structure of the activated voltage sensor

S4 segments in VSDs consist of repeated motifs of a positively charged residue, usually arginine, followed by two hydrophobic residues<sup>5–7</sup>. The R2 and R3 ‘gating charges’ in NavAb are positioned to interact with a conserved extracellular negative-charge cluster (ENC; Fig. 1b), whereas the R4 gating charge interacts with a conserved intracellular negative-charge cluster (INC; Fig. 1b). These structural features, in conjunction with disulphide-locking experiments<sup>16,17</sup>, indicate that the VSDs are in an activated conformation. These ion-pair interactions are expected to stabilize and catalyse S4 movement in the membrane electric field<sup>7,18,19</sup>.

<sup>1</sup>Department of Pharmacology, University of Washington, Seattle, Washington 98195, USA. <sup>2</sup>Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA.



**Figure 1 | Structure of NavAb and the activated VSD.** **a**, Structural elements in NavAb. One subunit is highlighted (1–6, transmembrane segments S1–S6). The nearest VSD has been removed for clarity. **b**, Side and top views of the VSD illustrating the ENC (red), INC (red), HCS (green), residues of the S1N helix (cyan) and phenylalanines of the S2–S3 loop (purple). S4 segment and gating

charges (R1–R4) are in yellow. **c–e**, Hydrogen bonding of gating charges, dotted lines ( $<3.5$  Å).  $F_o - F_c$  omit maps are contoured over E96 and R1–R4 at 1, 1, 1.5, 2.5 and 1.75 $\sigma$ , respectively. **f**, S3–S4 loop. Coloured according to crystallographic temperature factors of the main chain (blue  $<50$  Å<sup>2</sup> to red  $>150$  Å<sup>2</sup>). An  $F_o - F_c$  omit map is contoured at 1.5 $\sigma$  (grey) and 2.5 $\sigma$  (pink).

Highly conserved Arg 63 in the S2 segment also interacts with R4 and the INC (Fig. 1e), which may stabilize the INC and modulate its electrostatics<sup>20</sup>. NavAb has a spectrum of additional gating charge interactions. R1 interacts with Glu 96, R2 forms a hydrogen bond with the backbone carbonyl of Val 89 in S3, and R3 forms hydrogen bonds with Asn 25 and Met 29 in S1, and Ser 87 in S3 (Fig. 1c–e). This conserved network of hydrogen bonds (Supplementary Fig. 7a) should complement exchange of ion-pair partners and provide a low-energy pathway for S4 movement. The R2–backbone interaction would escape detection in mutagenesis experiments (Fig. 1c) and could have unrecognized significance in the passage of gating charges through the gating pore (Fig. 1b).

The S4 segment in NavAb forms a  $3_{10}$ -helix from R1 to R4. This conformation places all four gating charges in a straight line on one side of S4 (Fig. 1b), such that they could move linearly through the central portion of the gating pore, rather than with a spiral motion<sup>7,17–19</sup>. The S3 segment is a straight  $\alpha$ -helix, and the S3–S4 loop has a dynamic connection to S4 (Fig. 1f). The lack of structural rigidity within the S3–S4 loop (Fig. 1f) indicates that it could move relatively freely in response to large S4 movements during gating.

Our structural analysis reveals further that the S1N helix and S2–S3 loop shield the intracellular surface of the VSD (Fig. 1b and Supplementary Fig. 8). The S2–S3 loop is conserved among VGICs, and two prominent Phe side chains probably stabilize the VSD in the membrane during gating transitions (Fig. 1b and Supplementary Figs 7 and 8)<sup>9</sup>. The S1N-to-S3 region may behave as a modular unit during activation. In contrast to the sheltered intracellular surface of the VSD, a large aqueous cleft extends  $\sim 10$  Å from the extracellular surface into the membrane region above the hydrophobic constriction site (HCS; Fig. 1b). The HCS contains highly conserved residues (Ile 22, Phe 56 and Val 84; Supplementary Fig. 7) that seal the VSD against ion leakage

during S4 movement (Fig. 1b). The NavAb VSD therefore illustrates two important concepts from structure–function studies of Na<sub>v</sub> channels: a large external vestibule accessible to hydrophilic reagents; and a focused membrane electric field over the intracellular half of the VSD<sup>6,7</sup>.

Despite their separation over one billion years of evolution, the VSDs of NavAb and K<sub>v</sub>1.2 show highly similar conformations (Supplementary Fig. 8a). R4 of NavAb is in an equivalent position to K5 in K<sub>v</sub>1.2 (Supplementary Fig. 8a), the most outward location of K5 during voltage-sensor activation<sup>20</sup>. This observation implies that the NavAb and K<sub>v</sub>1.2 VSDs are both activated.

### The NavAb activation gate is closed

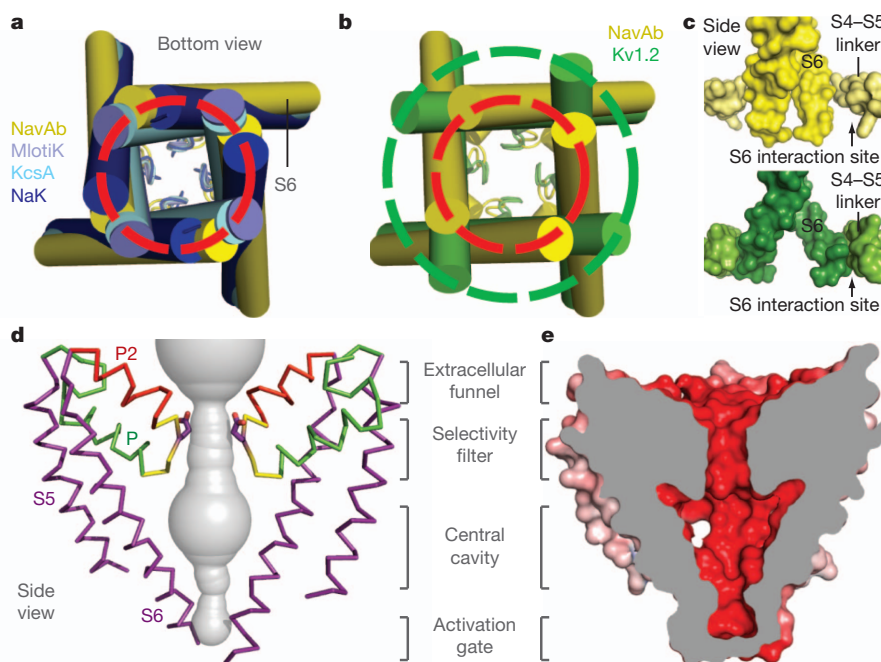
The pore of NavAb is closed, providing the first view of a closed pore in a VGIC (Fig. 2a and Supplementary Fig. 3). Met 221 completely occludes the ion conduction pathway (Supplementary Fig. 4c). The S6 helices of NavAb superimpose well with other closed-pore structures and are distinct from the open-pore K<sub>v</sub>1.2 structure (Fig. 2a, b). A subtle iris-like dilation of the activation gate may be sufficient to open the pore, and the surrounding cuff of S4–S5 linkers may prevent larger pore opening (Fig. 2a–c).

It is surprising to have a closed pore in a VGIC with activated voltage sensors at 0 mV. Our NavAb structures were obtained by introducing a Cys at two locations near the intracellular end of S6 (Ile217Cys or Met221Cys). Evidently, these substitutions allowed us to trap the NavAb channel in the pre-open state previously invoked in kinetic models of VGIC gating (Supplementary Discussion)<sup>21–23</sup>.

### Architecture of the pore and selectivity filter

VGICs are selective for specific cations yet conduct these ions at nearly the rate of free diffusion<sup>2</sup>. Our NavAb structure uncovers a





**Figure 2 | NavAb pore module.** **a**, Pore-lining S6 helices of NavAb (yellow) and the closed MlotiK (PDB code 3BEH), KcsA (PDB code 1K4C) and NaK (PDB code 2AHY) channels. C $\alpha$  locations of Met 221 define a common radius for the closed activation gate (red circle). **b**, Comparison of S6 helices of NavAb and Kv1.2/2.1 (PDB code 2R9R). Dashed circle in red indicates radius of C $\alpha$

atoms of Met 221 in NavAb. **c**, Site for interaction of S6 with S4–S5 linkers (top, NavAb; bottom, Kv1.2/2.1). **d**, Architecture of the NavAb pore. Glu 177 side chains (purple sticks); pore volume is shown in grey. **e**, Electrostatic potential coloured from  $-10$  to  $10$  kT (red to blue).

basis for the selectivity and high conductance of Na $_V$  channels. The NavAb pore module consists of an outer funnel-like vestibule, a selectivity filter, a central cavity and an intracellular activation gate (Fig. 2d and Supplementary Fig. 4b). The large central cavity in NavAb could easily accommodate a Na $^+$  ion with its first hydration shell and would present a hydrophobic surface over which ions should rapidly diffuse (Fig. 2e and Supplementary Figs 1 and 9). The pore (P)-helices are positioned to stabilize cations in the central cavity through helical–dipole interactions (Fig. 2d and Supplementary Fig. 4b), as suggested for K $^+$  channels<sup>24,25</sup>. Notably, a second pore-helix (P2-helix) forms an extracellular funnel in NavAb (Fig. 2d). This unique P2-helix is not seen in K $^+$  channels and may represent a conserved structural element in the outer vestibule of Na $_V$  and Ca $_V$  channels.

The ion conduction pathway in NavAb is strongly electronegative and the selectivity filter forms the narrowest constriction near the extracellular side of the membrane (Figs 2d, e, 3 and Supplementary Fig. 9). Classic permeation studies suggested a high-field-strength anionic site with dimensions of  $\sim 3.1 \times 5.1$  Å for the selectivity filter in Na $_V$  channels<sup>26,27</sup> and  $5.5 \times 5.5$  Å in Ca $_V$  channels<sup>28</sup>. Mutagenesis studies implicated Glu side chains as key determinants of ion selectivity in these channels<sup>29–33</sup>. In NavAb, the four Glu 177 side chains form a  $\sim 6.5 \times 6.5$  Å scaffold with an orifice of  $\sim 4.6 \times 4.6$  Å defined by van der Waals surfaces (Fig. 3a and Supplementary Fig. 9d). Remarkably, Glu 177 aligns with Glu residues that determine ion selectivity in Na $_V$  and Ca $_V$  channels (Fig. 3e).

The Glu 177 side chains of NavAb are supported by an elaborate architecture (Supplementary Figs 10 and 11). The P-helix ends with the conserved Thr 175, which accepts a hydrogen bond (3.0 Å) from the conserved Trp 179 of a neighbouring subunit (Fig. 3a). This landmark interaction staples together adjacent subunits at the selectivity filter. The residues between Thr 175 and Trp 179 form a tight turn and expose backbone carbonyls of Thr 175 and Leu 176 to conducted ions (Fig. 3b). The Glu 177 side chains form hydrogen bonds with the backbone amides of Ser 180 (2.6 Å) and Met 181 (3.1 Å) from the P2-helix (Fig. 3b and Supplementary Fig. 10). An extensive network of additional interactions (Supplementary Fig. 10), including hydrogen

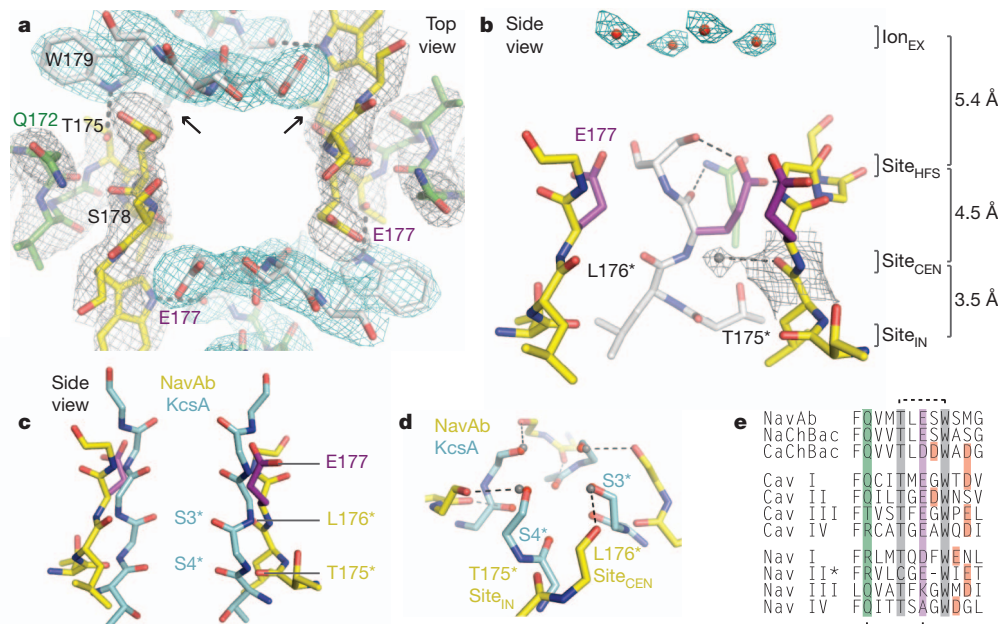
bonds between Glu 172 from the P-helix and the carbonyl of Glu 177 (Fig. 3a, b), further stabilizes the selectivity filter. Owing to the dimer-of-dimers arrangement, the Glu 177 and Ser 178 side chains of NavAb are in two slightly different environments (Fig. 3a and Supplementary Fig. 11), consistent with functional nonequivalence of the corresponding glutamates in Ca $_V$  channels<sup>31–33</sup>.

In agreement with the low affinity of Na $_V$  channels for permeant ions ( $K_d$  for Na $^+ > 350$  mM<sup>34</sup>), no extra density was observed beside the Glu 177 side chains. Instead, strong electron densities were found above Glu 177 at a distance of  $> 4$  Å. These densities probably represent cations or solvent molecules (Ion $_{EX}$ ; Fig. 3b) positioned above the selectivity filter by its intense electronegativity (Fig. 2e).

### Ion permeation and selectivity

NavAb represents a prototype for understanding Na $^+$  selectivity and permeation. Analysis of the pore radius indicates that a partially hydrated Na $^+$  ion can be accommodated at the high-field-strength site formed by the Glu 177 side chains (Site $_{HFS}$ ; Fig. 3a, b and Supplementary Fig. 9d). The much narrower K $^+$ -channel filter can fit inside the NavAb selectivity filter (Fig. 3c). Careful inspection of the electron density indicates four well-bound water molecules 2.5 Å from the Leu 176 carbonyls (Site $_{CEN}$ ; Fig. 3b). Remarkably, these four water molecules occupy the same positions as the site 3 carbonyls from K $^+$  channels (Fig. 3c, d)<sup>35</sup>. A distance of 2.5 Å is also found between the backbone carbonyls of Thr 175 from NavAb and the site 4 carbonyls of K $^+$  channels (Fig. 3c, d)<sup>35</sup>. Analogous to other Na $^+$  complexes (Supplementary Fig. 12)<sup>36–40</sup>, a Na $^+$  ion surrounded by a square array of four water molecules could interact with the backbone carbonyls of Leu 176 (Site $_{CEN}$ ) or Thr 175 (Site $_{IN}$ ) (Fig. 3d and Supplementary Fig. 12). Therefore, unlike K $^+$  channels, the NavAb selectivity filter seems to select and conduct Na $^+$  ions in a mostly hydrated form.

The NavAb structure fits closely with Hille's single-ion pore model for Na $_V$  channels, in which a high-field-strength anion partially dehydrates the permeating ion<sup>2,34</sup>. According to Eisenman's theory<sup>41</sup>, a Na $^+$  ion would approach the Site $_{HFS}$  more closely than the larger



**Figure 3 | Structure of the NavAb selectivity filter.** **a**, Top view of the selectivity filter. Symmetry-related molecules are coloured white and yellow; P-helix residues are coloured green. Hydrogen bonds between Thr 175 and Trp 179 are indicated by grey dashes. Electron densities from  $F_o - F_c$  omit maps are contoured at  $4.0\sigma$  (blue and grey) and subtle differences can be appreciated (small arrows). **b**, Side view of the selectivity filter. Glu 177 (purple) interactions with Gln 172, Ser 178 and the backbone of Ser 180 are shown in the far subunit.  $F_o - F_c$  omit map,  $4.75\sigma$  (blue); putative cations or water molecules (red spheres,  $\text{Ion}_{\text{EX}}$ ). Electron density around Leu 176 (grey;  $F_o - F_c$  omit map

$\text{K}^+$  ion, allowing more efficient removal of water and faster permeation (Fig. 3a, b)<sup>34</sup>. A  $\text{Na}^+$  ion could fit in-plane between the Glu 177 side chains, with one side chain coordinating the  $\text{Na}^+$  ion directly and neighbouring Glu 177 side chains acting as hydrogen bond acceptors for two in-plane water molecules<sup>26,27,34</sup>. With two additional waters remaining axial to the ion, this arrangement would approximate trigonal bipyramidal coordination<sup>38</sup>. Because only one Glu 177 side chain engages the permeating ion directly, this transient complex would be inherently asymmetric. When the permeating ion escapes Site<sub>HFS</sub>, full rehydration would occur along the water-lined sites formed by the backbone carbonyls of Leu 176 (Site<sub>CEN</sub>) and Thr 175 (Site<sub>IN</sub>; Fig. 3b, d and Supplementary Fig. 12). Free diffusion then allows the hydrated  $\text{Na}^+$  ion to enter the central cavity and move through the open activation gate into the cytoplasm<sup>34</sup>. The selectivity-filter structure of NavAb concentrates barriers to ion flow into  $\sim 5 \text{ \AA}$  (Fig. 3b and Supplementary Fig. 9d), which should promote high flux rates<sup>34</sup>. This permeation mechanism probably reflects the high free energy of  $\text{Na}^+$  hydration, where further removal of solvating waters would present too high an energy barrier. In sharp contrast,  $\text{K}^+$ -selective channels conduct nearly fully dehydrated  $\text{K}^+$  ions through direct interactions with backbone carbonyls in a long, narrow, multi-ion pore<sup>35,36</sup>. The architectures of the selectivity filters of vertebrate  $\text{Na}_v$  and  $\text{Ca}_v$  channels probably resemble NavAb, and amino acid substitutions within this structural framework must impart  $\text{Na}^+$  versus  $\text{Ca}^{2+}$  selectivity (Supplementary Discussion)<sup>13,29–33</sup>.

### Interaction sites of pore blockers

NavAb provides a foundation to interpret pharmacological mechanisms. From the extracellular side, the Glu 177 side chains of NavAb represent the blocking site in  $\text{Na}_v$  channels for protons and guanidinium moieties of tetrodotoxin and saxitoxin<sup>2,42</sup>, as well as the site where divalent cations and protons bind and block  $\text{Ca}_v$  channels (Fig. 3)<sup>31–33</sup>. From the intracellular side, local anaesthetics, antiarrhythmics and

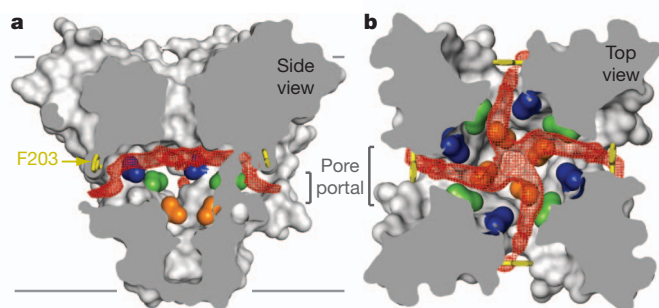
antiepileptic drugs block  $\text{Na}_v$  and  $\text{Ca}_v$  channels<sup>2,4</sup> by entering through the open intracellular mouth of the pore and binding to an overlapping receptor site on the S6 segments<sup>43–45</sup>. Alignment of NavAb S6 segments with vertebrate  $\text{Na}_v$  and  $\text{Ca}_v$  channels reveals a high degree of sequence similarity (Supplementary Fig. 7b), and drug molecules could easily fit into the large central cavity (Fig. 2e and Supplementary Fig. 9). Use-dependent block is enhanced by repetitive opening of the pore to provide drug access<sup>2,46</sup>, and the local anaesthetic etidocaine is an open-channel blocker of NaChBac<sup>12</sup>. The tight seal observed at the intracellular activation gate in NavAb illustrates why pore opening is required for access of large or hydrophilic drugs to the S6 receptor site (Fig. 2 and Supplementary Fig. 4c).

### Fenestrations provide hydrophobic access to pore

Membrane lipids modulate the structure and function of VGICs<sup>8,9,47,48</sup>. However, NavAb presents a completely unexpected type of lipid interaction that has profound implications. The NavAb central cavity reveals four lateral openings leading from the membrane to the lumen of the closed pore (Fig. 4). These fenestrations measure  $\sim 8 \times 10 \text{ \AA}$ , and could become larger depending upon nearby side-chain conformations (Phe 203; Fig. 4). Lipids penetrate through these side portals and lie deep within the central cavity, occluding the ion conduction pathway in NavAb (Fig. 4, red). Because acyl-chain-containing detergents were never used in the preparation of NavAb crystals, these electron densities are assigned as acyl chains of membrane phospholipids. Similar fenestrations were not observed in the open-pore structure of  $\text{K}_v1.2$  (refs 8, 9), raising the possibility that these lipid chains withdraw and the fenestrations close in the open state.

The lateral pore fenestrations in NavAb lead directly to the drug-binding sites within the central cavity and abut residues that are important for drug binding in  $\text{Na}_v$  and  $\text{Ca}_v$  channels (Fig. 4 and Supplementary Fig. 7b)<sup>43,44</sup>. These NavAb portals appear compatible





**Figure 4 | Membrane access to the central cavity in NavAb.** **a**, Side view through the pore module illustrating fenestrations (portals) and hydrophobic access to central cavity. Phe 203 side chains are shown as yellow sticks. Surface representations of NavAb residues aligning with those implicated in drug binding and block: Thr 206, blue; Met 209, green; Val 213, orange. Membrane boundaries, grey lines. Electron density from an  $F_o - F_c$  omit map is contoured at  $2.0\sigma$ . **b**, Top view sectioned below the selectivity filter, coloured as in **a**.

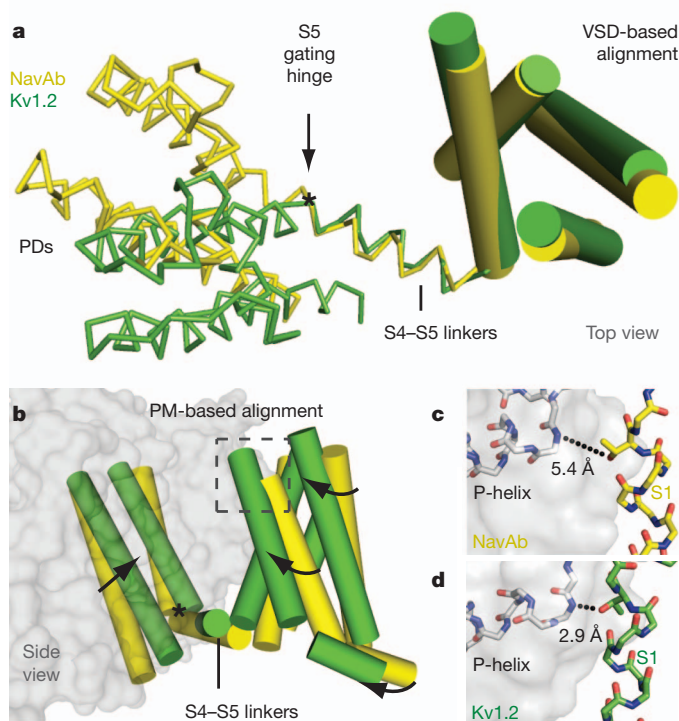
with the passage of small neutral or hydrophobic drugs such as phenytoin<sup>49</sup> and benzocaine<sup>46</sup>, which can gain access to their receptor site in closed channels<sup>2,46</sup>. We propose that pore fenestrations may be directly involved in voltage-dependent drug block according to the ‘modulated receptor model’<sup>46</sup>. Our findings highlight the potential for lipids and other hydrophobic molecules to influence the function of ion channels from the lipid phase of the membrane.

### Structural basis for central pore gating

The domain-swapped arrangement of the VSD around the pore allows the S4–S5 linker to couple S4 movements to activation of VGICs (Fig. 1a)<sup>9</sup>. Kinetic models indicate that all four voltage sensors activate and then the central pore opens in a concerted transition<sup>21–23</sup>. An essential element of this gating model is a state in which all four VSDs have activated but the pore remains closed<sup>21–23</sup>. It is likely that we have captured this pre-open state in our crystals (Supplementary Discussion). NavAb therefore provides a unique opportunity to consider the structural basis for coupling of VSD activation to pore opening.

When activated VSDs of NavAb and Kv1.2 are overlaid (Supplementary Fig. 8a), the S4–S5 linkers superimpose precisely, but the pore domains diverge at the foot of S5 (Fig. 5a). Superposition of the pore domains demonstrates an equivalent displacement of the VSDs (Supplementary Fig. 13). These comparisons lead to a working model for pore opening. First, during activation, the S4–S5 linker and VSD move together as a modular unit (Fig. 5a). Second, a single molecular hinge at the base of S5 mediates the closed-to-open pore transition (Fig. 5a, b). Third, tight structural coupling is maintained between the S5 and S6 segments (Supplementary Fig. 13a). This model suggests that rotation of the VSD and S4–S5 linker as a structural unit pulls the S5–S6 helices outward to open the pore (Fig. 5b and Supplementary Fig. 13b). Because of their tight structural coupling, displacement of the S5–S6 segments from one subunit forces the neighbouring subunits to move similarly, leading to concerted pore opening. During this transition, the amphipathic S4–S5 linker pivots along the plane of the membrane interface (Fig. 5b and Supplementary Figs 7 and 13b). In contrast to Kv1.2, the S6 helices in NavAb have not fully engaged their interaction site on the S4–S5 linker (Fig. 2c), in agreement with the pre-open state of NavAb. A rolling motion of the VSDs around the pore produces displacements up to  $\sim 10$  Å at the intracellular side (Fig. 5b and Supplementary Fig. 13b), which may influence movements of the S1N helix and the conserved S2–S3 loop.

In NavAb, a  $3_{10}$ -helix extends from R1 to R4 (Fig. 1b). In Kv1.2, a  $3_{10}$ -helix encompasses R3 to K5 (equivalent to NavAb R2 to R4), but



**Figure 5 | Model for activation gate opening.** **a**, Superposition of NavAb and Kv1.2/2.1 on the basis of their VSDs (cylinders). PDs, pore domains. **b**, Superposition of NavAb and Kv1.2/2.1 tetrameric pore modules (PM) viewed from the membrane. S5 gating hinge is indicated with an asterisk. Dashed square is enlarged in panels **c** and **d**. **c**, **d**, S1 interaction with P-helix. The distance from the S1 Thr to the P-helix of the neighbouring subunit is 2.9 Å in Kv1.2/2.1, but  $>4.5$  Å in NavAb.

the remaining S4 segment is  $\alpha$ -helical<sup>9</sup>. Conceivably, energy derived from voltage-driven translocation of S4 may be stored in the higher-energy  $3_{10}$ -helix, and then released to help drive pore opening. The VSDs in Kv1.2 are displaced outward ( $\sim 2$  Å) compared to the pre-open NavAb structure (Fig. 5b), which could account for the small gating current associated with concerted pore opening<sup>6</sup>. At the extracellular side of the VSD, an S1 threonine residue hydrogen bonds (2.9 Å) with the P-helix of a neighbouring subunit in Kv1.2 (Fig. 5d), providing a conserved contact point that allows the VSD to perform mechanical work on the pore<sup>50</sup>. The equivalent S1 threonine in NavAb has not yet engaged the P-helix (Fig. 5c). This interaction may therefore represent an essential step in activation gating that has not yet occurred in the pre-open state of NavAb.

### Conclusion

The structure of NavAb provides key insights into the molecular basis of voltage sensing, ion conductance and voltage-dependent gating in a historic class of ion channels<sup>1,2</sup>. A new network of interactions within the VSD appears well positioned to catalyse gating charge movements during activation. Our model for electromechanical coupling reveals a rolling motion of the VSD and its connecting S4–S5 linker around the pore. The NavAb selectivity filter illustrates the basis for selective  $\text{Na}^+$  conductance through a water-lined pore featuring a high-field-strength anionic site. Lastly, hydrophobic access from the membrane phase has been uncovered as a potentially important pathway for drug binding and modulation of VGICs.

### METHODS SUMMARY

NavAb was expressed in insect cells and purified using anti-Flag resin and size-exclusion chromatography, reconstituted into DMPC:CHAPSO bicelles, and crystallized over an ammonium sulphate solution containing 0.1 M Na-citrate,



pH 4.75. Cysteine mutants were complexed with mercury to obtain initial experimental phases. A single anomalous dispersion (SAD) data set from a mercury-free SeMet-substituted protein crystal expedited model building. Standard crystallographic refinement procedures and structural analyses were performed. Electrophysiological experiments on NavAb were performed in tsA-201 cells using standard protocols.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 16 January; accepted 27 May 2011.**

**Published online 10 July 2011.**

- Hodgkin, A. L. & Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol. (Lond.)* **117**, 500–544 (1952).
- Hille, B. *Ion Channels of Excitable Membranes* 3rd edn (Sinauer Associates, 2001).
- Ryan, D. P. & Ptacek, L. J. Episodic neurological channelopathies. *Neuron* **68**, 282–292 (2010).
- Catterall, W. A. Common modes of drug action on Na<sup>+</sup> channels: local anesthetics, antiarrhythmics and anticonvulsants. *Trends Pharmacol. Sci.* **8**, 57–65 (1987).
- Yu, F. H. & Catterall, W. A. The VGL-channome: a protein superfamily specialized for electrical signaling and ionic homeostasis. *Sci. STKE* **2004**, re15 (2004).
- Bezanilla, F. The action potential: from voltage-gated conductances to molecular structures. *Biol. Res.* **39**, 425–435 (2006).
- Catterall, W. A. Ion channel voltage sensors: structure, function, and pathophysiology. *Neuron* **67**, 915–928 (2010).
- Long, S. B., Campbell, E. B. & MacKinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K<sup>+</sup> channel. *Science* **309**, 897–903 (2005).
- Long, S. B., Tao, X., Campbell, E. B. & MacKinnon, R. Atomic structure of a voltage-dependent K<sup>+</sup> channel in a lipid membrane-like environment. *Nature* **450**, 376–382 (2007).
- Ren, D. *et al.* A prokaryotic voltage-gated sodium channel. *Science* **294**, 2372–2375 (2001).
- Koishi, R. *et al.* A superfamily of voltage-gated sodium channels in bacteria. *J. Biol. Chem.* **279**, 9532–9538 (2004).
- Zhao, Y., Scheuer, T. & Catterall, W. A. Reversed voltage-dependent gating of a bacterial sodium channel with proline substitutions in the S6 transmembrane segment. *Proc. Natl Acad. Sci. USA* **101**, 17873–17878 (2004).
- Yue, L., Navarro, B., Ren, D., Ramos, A. & Clapham, D. E. The cation selectivity filter of the bacterial sodium channel, NaChBac. *J. Gen. Physiol.* **120**, 845–853 (2002).
- Curtis, B. M. & Catterall, W. A. Reconstitution of the voltage-sensitive calcium channel purified from skeletal muscle transverse tubules. *Biochemistry* **25**, 3077–3083 (1986).
- Feller, D. J., Talvenheimo, J. A. & Catterall, W. A. The sodium channel from rat brain. Reconstitution of voltage-dependent scorpion toxin binding in vesicles of defined lipid composition. *J. Biol. Chem.* **260**, 11542–11547 (1985).
- DeCaen, P. G., Yarov-Yarovoy, V., Zhao, Y., Scheuer, T. & Catterall, W. A. Disulfide locking a sodium channel voltage sensor reveals ion pair formation during activation. *Proc. Natl Acad. Sci. USA* **105**, 15142–15147 (2008).
- DeCaen, P. G., Yarov-Yarovoy, V., Sharp, E. M., Scheuer, T. & Catterall, W. A. Sequential formation of ion pairs during activation of a sodium channel voltage sensor. *Proc. Natl Acad. Sci. USA* **106**, 22498–22503 (2009).
- Catterall, W. A. Molecular properties of voltage-sensitive sodium channels. *Annu. Rev. Biochem.* **55**, 953–985 (1986).
- Yarov-Yarovoy, V., Baker, D. & Catterall, W. A. Voltage sensor conformations in the open and closed states in ROSETTA structural models of K<sup>+</sup> channels. *Proc. Natl Acad. Sci. USA* **103**, 7292–7297 (2006).
- Tao, X., Lee, A., Limapichat, W., Dougherty, D. A. & MacKinnon, R. A gating charge transfer center in voltage sensors. *Science* **328**, 67–73 (2010).
- Zagotta, W. N., Hoshi, T. & Aldrich, R. W. Shaker potassium channel gating. III: Evaluation of kinetic models for activation. *J. Gen. Physiol.* **103**, 321–362 (1994).
- Kuzmenkin, A., Bezanilla, F. & Correa, A. M. Gating of the bacterial sodium channel, NaChBac: voltage-dependent charge movement and gating currents. *J. Gen. Physiol.* **124**, 349–356 (2004).
- Zhao, Y., Yarov-Yarovoy, V., Scheuer, T. & Catterall, W. A. A gating hinge in Na<sup>+</sup> channels; a molecular switch for electrical signaling. *Neuron* **41**, 859–865 (2004).
- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K<sup>+</sup> conduction and selectivity. *Science* **280**, 69–77 (1998).
- Jogini, V. & Roux, B. Electrostatics of the intracellular vestibule of K<sup>+</sup> channels. *J. Mol. Biol.* **354**, 272–288 (2005).
- Hille, B. The permeability of the sodium channel to organic cations in myelinated nerve. *J. Gen. Physiol.* **58**, 599–619 (1971).
- Hille, B. The permeability of the sodium channel to metal cations in myelinated nerve. *J. Gen. Physiol.* **59**, 637–658 (1972).
- McCleskey, E. W. & Almers, W. The Ca channel in skeletal muscle is a large pore. *Proc. Natl Acad. Sci. USA* **82**, 7149–7153 (1985).
- Heinemann, S. H., Terlau, H., Stuhmer, W., Imoto, K. & Numa, S. Calcium channel characteristics conferred on the sodium channel by single mutations. *Nature* **356**, 441–443 (1992).
- Favre, I., Moczydlowski, E. & Schild, L. On the structural basis for ionic selectivity among Na<sup>+</sup>, K<sup>+</sup>, and Ca<sup>2+</sup> in the voltage-gated sodium channel. *Biophys. J.* **71**, 3110–3125 (1996).
- Yang, J., Ellinor, P. T., Sather, W. A., Zhang, J. F. & Tsien, R. W. Molecular determinants of Ca<sup>2+</sup> selectivity and ion permeation in L-type Ca<sup>2+</sup> channels. *Nature* **366**, 158–161 (1993).
- Ellinor, P. T., Yang, J., Sather, W. A., Zhang, J. F. & Tsien, R. W. Ca<sup>2+</sup> channel selectivity at a single locus for high-affinity Ca<sup>2+</sup> interactions. *Neuron* **15**, 1121–1132 (1995).
- Chen, X. H., Bezprozvanny, I. & Tsien, R. W. Molecular basis of proton block of L-type Ca<sup>2+</sup> channels. *J. Gen. Physiol.* **108**, 363–374 (1996).
- Hille, B. Ionic selectivity, saturation, and block in sodium channels. A four-barrier model. *J. Gen. Physiol.* **66**, 535–560 (1975).
- Morais-Cabral, J. H., Zhou, Y. & MacKinnon, R. Energetic optimization of ion conduction rate by the K<sup>+</sup> selectivity filter. *Nature* **414**, 37–42 (2001).
- Ye, S., Li, Y. & Jiang, Y. Novel insights into K<sup>+</sup> selectivity from high-resolution structures of an open K<sup>+</sup> channel pore. *Nature Struct. Mol. Biol.* **17**, 1019–1023 (2010).
- Alam, A. & Jiang, Y. Structural analysis of ion selectivity in the NaK channel. *Nature Struct. Mol. Biol.* **16**, 35–41 (2009).
- Doi, M. *et al.* Caged and clustered structures of endothelin inhibitor BQ123, cyclo-D-Trp-D-Asp<sup>+</sup>-Pro-D-Val-Leu<sup>+</sup>-Na<sup>+</sup>, forming five and six coordination bonds between sodium ions and peptides. *Acta Crystallogr. D* **57**, 628–634 (2001).
- Harding, M. M. Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium. *Acta Crystallogr. D* **58**, 872–874 (2002).
- Phillips, K., Dauter, Z., Murchie, A. I., Lilley, D. M. & Luisi, B. The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution. *J. Mol. Biol.* **273**, 171–182 (1997).
- Eisenman, G. & Horn, R. Ionic selectivity revisited: the role of kinetic and equilibrium processes in ion permeation through channels. *J. Membr. Biol.* **76**, 197–225 (1983).
- Noda, M., Suzuki, H., Numa, S. & Stuhmer, W. A single point mutation confers tetrodotoxin and saxitoxin insensitivity on the sodium channel II. *FEBS Lett.* **259**, 213–216 (1989).
- Hockerman, G. H., Peterson, B. Z., Johnson, B. D. & Catterall, W. A. Molecular determinants of drug binding and action on L-type calcium channels. *Annu. Rev. Pharmacol. Toxicol.* **37**, 361–396 (1997).
- Ragsdale, D. S., McPhee, J. C., Scheuer, T. & Catterall, W. A. Molecular determinants of state-dependent block of Na<sup>+</sup> channels by local anesthetics. *Science* **265**, 1724–1728 (1994).
- Ragsdale, D. S., McPhee, J. C., Scheuer, T. & Catterall, W. A. Common molecular determinants of local anesthetic, antiarrhythmic, and anticonvulsant block of voltage-gated Na<sup>+</sup> channels. *Proc. Natl Acad. Sci. USA* **93**, 9270–9275 (1996).
- Hille, B. Local anesthetics: hydrophilic and hydrophobic pathways for the drug-receptor reaction. *J. Gen. Physiol.* **69**, 497–515 (1977).
- Oliver, D. *et al.* Functional conversion between A-type and delayed rectifier K<sup>+</sup> channels by membrane lipids. *Science* **304**, 265–270 (2004).
- Delmas, P., Coste, B., Gamper, N. & Shapiro, M. S. Phosphoinositide lipid second messengers: new paradigms for calcium channel modulation. *Neuron* **47**, 179–182 (2005).
- Morello, R. S., Begenisich, T. & Yeh, J. Z. Determination of the active form of phenytoin. *J. Pharmacol. Exp. Ther.* **230**, 156–161 (1984).
- Lee, S. Y., Banerjee, A. & MacKinnon, R. Two separate interfaces between the voltage sensor and pore are required for the function of voltage-dependent K<sup>+</sup> channels. *PLoS Biol.* **7**, e47 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank B. Hille for comments on a draft of the manuscript and members of the N.Z. and W.A.C. groups for their support throughout this project. We are grateful to investigators who provided genomic DNA and the beamline staff at the Advanced Light Source (BL8.2.1 and BL8.2.2) for their assistance during data collection. J.P. acknowledges support from a Canadian Institutes of Health Research fellowship and the encouragement of E. Payandeh. This work was supported by grants from the National Institutes of Health (R01 NS15751 and U01 NS058039 to W.A.C.) and by the Howard Hughes Medical Institute (N.Z.).

**Author Contributions** N.Z. and W.A.C. are co-senior authors. J.P., N.Z. and W.A.C. conceived and J.P. conducted the protein purification and crystallization experiments. J.P. and N.Z. determined and analysed the structures of NavAb. J.P. and T.S. performed functional studies of NavAb. J.P., N.Z. and W.A.C. wrote the manuscript.

**Author Information** Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 3RVY, 3RVZ and 3RWO. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to N.Z. (nzheng@uw.edu) and W.A.C. (wcatt@uw.edu).

## METHODS

**Protein expression and purification.** After exploring traditional expression approaches in *Escherichia coli*<sup>51</sup>, the NavAb channel from *A. butzleri* was cloned into the pFASTBac-Dual vector behind the polyhedron promoter using the BamHI and NotI restriction sites preceded by an N-terminal Flag tag. Recombinant baculovirus were generated using the Bac-to-Bac system (Invitrogen) and insect cells were infected for large-scale protein production. Cells were harvested 72 h post-infection and resuspended in 50 mM Tris pH 8.0, 200 mM NaCl (Buffer A) supplemented with protease inhibitors and DNase. After sonication, digitonin (EMD Biosciences) was added to 1% and solubilization was carried out for 1–2 h at 4 °C. After centrifugation, clarified supernatant was gently agitated with anti-Flag M2-agarose resin (Sigma) pre-equilibrated with Buffer B (Buffer A supplemented with 0.12% digitonin) for 1–2 h at 4 °C. Flag resin was collected in a column by gravity flow, washed with ten column volumes of Buffer B, and eluted with two column volumes of Buffer B supplemented with 0.1 mg ml<sup>-1</sup> Flag peptide. The eluate was passed over a Superdex 200 column (GE Healthcare) in 10 mM Tris pH 8, 100 mM NaCl and 0.12% digitonin and peak fractions containing NavAb were concentrated using a Vivaspinn (30K MWKO) centrifugal device. Site-directed mutagenesis was performed using the standard QuikChange protocol (Stratagene) and all constructs were confirmed by DNA sequencing. Selenomethionine-labelled proteins were expressed using established protocols<sup>52</sup>, except cells were washed and starved for methionine at 8 h after infection, followed by SeMet (Anatrace) supplementation at 12 h after infection. SeMet-labelled proteins were purified as described earlier. **Heavy atom screening and labelling.** During our efforts to identify useful derivatives for crystallographic phasing, we ultimately turned to the fluorescence detection of heavy atom labelling (FD-HAL) method<sup>53</sup>. Over thirty NavAb single-site cysteine mutations were rapidly screened using the FD-HAL method, and many of these mutant proteins were subsequently crystallized, presumably as covalent mercury-channel complexes. The NavAb(Ile217Cys) and NavAb(Met221Cys) mutants that yielded useful single anomalous dispersion (SAD) data sets were prepared as follows: proteins were purified as described earlier and concentrated to ~1 mg ml<sup>-1</sup>; HgCl<sub>2</sub> was added to a final concentration of 10 mM and the mixture was incubated at room temperature (22 °C) for 2 h. The protein buffer was subsequently exchanged (into mercury-free buffer) through five rounds of concentration and dilution using Vivaspinn (30K MWKO) centrifugal devices. Following structure determination, it became apparent that Met 221 lines the narrowest portion of the closed NavAb pore.

**NavAb crystallization and data collection.** Before crystallization, NavAb was concentrated to ~20 mg ml<sup>-1</sup> and reconstituted into DMPC:CHAPSO (Anatrace) bicelles according to standard protocols<sup>54,55</sup>. The NavAb-bicelle preparation was mixed in a 1:1 ratio and setup in a hanging-drop vapour-diffusion format over a well solution containing 1.8–2.1 M ammonium sulphate, 100 mM Na-citrate pH 4.75. The mercury-free proteins, the mercury complexes, and the SeMet-labelled proteins all crystallized under essentially identical conditions. Crystals were typically passed through solutions containing 2 M ammonium sulphate, 100 mM Na-citrate pH 4.75 and 28% glucose (wt/v) in increments of ~6% glucose during harvesting. Crystals were plunged into liquid nitrogen and maintained at 100 K during all data collection procedures.

Over 1,000 crystals were screened and nearly 100 diffraction data sets were collected at a synchrotron radiation source (Advanced Light Source, BL8.2.1 and BL8.2.2). A SAD data set collected near the mercury absorption edge ( $\lambda = 1.005$  Å) from a mercury-containing complex of the NavAb(Ile217Cys) mutant was ultimately used to determine initial experimental phases. Our highest resolution SeMet SAD data set was collected near the selenium absorption edge ( $\lambda = 0.9795$  Å) from a mercury-free NavAb(Met221Cys) SeMet-labelled crystal. Subsequent native (that is, mercury-free) data sets were collected at standard wavelengths. Because the NavAb crystals were small (typically <0.15 mm × 0.15 mm × 0.15 mm), contained a high solvent content (~80%), were weakly diffracting, and radiation sensitive, special care was taken to minimize exposure times and to orient the crystals in order to maximize data completeness and quality.

**Structure determination and refinement.** X-ray diffraction data were integrated and scaled with the HKL2000 suite or DENZO/SCALEPACK<sup>56</sup> and, when required, further processed with the CCP4 package<sup>57</sup>. Experimental phases were determined using a 3.4 Å SAD data set from a Hg-containing NavAb(Ile217Cys) crystal. The SOLVE/RESOLVE<sup>58</sup> software were run in a standard setting and the first map, calculated at 3.7 Å, is shown in Supplementary Fig. 3. Ideal poly-alanine  $\alpha$ -helices were manually fitted into this map and the model was subsequently used in combined SAD-molecular replacement (MR) protocols within the Phenix software<sup>59</sup> using a 3.3 Å SAD data set obtained from a SeMet-labelled NavAb(Met221Cys) crystal. SAD-MR and MR-SAD-based maps were calculated and compared, allowing for complete register and amino acid assignment of the NavAb model. Higher-resolution native

data sets were ultimately obtained and phased by MR methods using the CNS suite<sup>60</sup> (although our best native NavAb(Met221Cys) data set is actually from a SeMet-containing crystal). Reiterative rounds of model building in O<sup>61</sup> were guided by inspection of omit maps and refinement with CNS<sup>60</sup> was performed with strict NCS-restraints, which were later relaxed during final rounds of refinement. Two strong densities (one per protein chain) assigned as solvent molecules (near the pore turret loop; not discussed in the main text) and all lipid molecules were added to the models at very late stages of refinement. Although trace amounts of digitonin are present in the crystallization condition, digitonin molecules were not readily observed in any electron density map. Refinement statistics, scaling statistics, and overall map quality were ultimately used to assign the NavAb space group as *I*222, although the data were found to closely mimic *I*422 ( $R_{\text{work}}/R_{\text{free}}$  stall at ~32% in *I*422).

**Structure analysis.** The geometry of NavAb structural models was assessed using PROCHECK<sup>62</sup>. The pore radius of NavAb was calculated using standard settings in the MOLE software<sup>63</sup>. Electrostatic surface calculations were performed with the APBS software<sup>64</sup>, calculated with 150 mM NaCl in the solvent. Structural alignments were performed using LSQMAN<sup>65</sup> and O<sup>61</sup>, where all channels were independently aligned onto NavAb based on the amino acid positions at the very beginning (that is, N-terminal portion) of their P-helices. The superposition of the atomic resolution Na<sup>+</sup>-complex structure<sup>40</sup> shown in Supplementary Fig. 12 was positioned manually, but the K<sup>+</sup>-channel and NaK-channel superpositions (Figs 2, 3, 5b and Supplementary Figs 12 and 13b) were obtained by simply aligning P-helices, as described earlier. All  $F_o - F_c$  omit maps shown throughout the main text and Supplementary Information have been calculated using standard settings and appropriate buffers in the CNS program<sup>60</sup>. The  $F_o - F_c$  omit map shown in Fig. 3b specifically derives from the 2.7 Å NavAb(Ile217Cys) data set and amino acids 170–183 were omitted from the calculation box. All structural figures were prepared with the PyMol software<sup>66</sup>.

**Electrophysiology.** NavAb was cloned into the CDM8 vector and transfected into tsA-201 cells (along with a CD8 marker construct) using standard protocols. Whole-cell currents were recorded with continuous perfusion of extracellular solution using an Axopatch 200 amplifier (Molecular Devices) with glass pipettes polished to 2–4 MΩ resistance. The intracellular pipette solution contained (in mM): 10 NaCl, 105 CsF, 20 TEA, 10 EGTA, 10 HEPES pH 7.4 (adjusted with CsOH). The extracellular Na<sup>+</sup> solution contained (in mM): 100 NaCl, 1 CaCl<sub>2</sub>, 1 MgCl<sub>2</sub>, 1 KCl, 50 TEA, 10 HEPES pH 7.4 (CsOH). For K-containing and Cs-containing extracellular solutions, NaCl was replaced with KCl or CsCl, respectively. The extracellular NMDG solution contained (in mM): 100 NMDG, 1 CaCl<sub>2</sub>, 1 MgCl<sub>2</sub>, 1 KCl, 50 TEA, 10 HEPES pH 7.4 (HCl) and the extracellular Ca<sup>2+</sup> solution contained (in mM): 75 CaCl<sub>2</sub>, 1 MgCl<sub>2</sub>, 1 KCl, 50 TEA, 10 HEPES pH 7.4 (CsOH). Voltage clamp pulses were generated and currents were recorded using Pulse software controlling an Instrutech ITC18 interface (HEKA). Data were analysed using Igor Pro 6.2 (WaveMetrics).

51. Koth, C. M. & Payandeh, J. Strategies for the cloning and expression of membrane proteins. *Adv. Protein Chem. Struct. Biol.* **76**, 43–86 (2009).
52. Cronin, C. N., Lim, K. B. & Rogers, J. Production of selenomethionyl-derivatized proteins in baculovirus-infected insect cells. *Protein Sci.* **16**, 2023–2029 (2007).
53. Chaptal, V. et al. Fluorescence Detection of Heavy Atom Labeling (FD-HAL): a rapid method for identifying covalently modified cysteine residues by phasing atoms. *J. Struct. Biol.* **171**, 82–87 (2010).
54. Faham, S. & Bowie, J. U. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J. Mol. Biol.* **316**, 1–6 (2002).
55. Faham, S. et al. Crystallization of bacteriorhodopsin from bicelle formulations at room temperature. *Protein Sci.* **14**, 836–840 (2005).
56. Otwinowski, Z. & Minor, W. *Processing of X-ray Diffraction Data Collected in Oscillation Mode* Vol. 276 (Academic, 1997).
57. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
58. Terwilliger, T. SOLVE and RESOLVE: automated structure solution and density modification. *Meth. Enzymol.* **374**, 22–37 (2003).
59. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
60. Brünger, A. T. et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
61. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
62. Laskowski, R. A., Moss, D. S. & Thornton, J. M. Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**, 1049–1067 (1993).
63. Petřek, M., Kosinova, P., Koca, J. & Otyepka, M. MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure* **15**, 1357–1363 (2007).
64. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
65. Kleywegt, G. J. Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr. D* **52**, 842–857 (1996).
66. DeLano, W. L. PyMOL molecular viewer (v. 1.2r3pre) (<http://www.pymol.org>) (2002).



# Crystal structure of the human centromeric nucleosome containing CENP-A

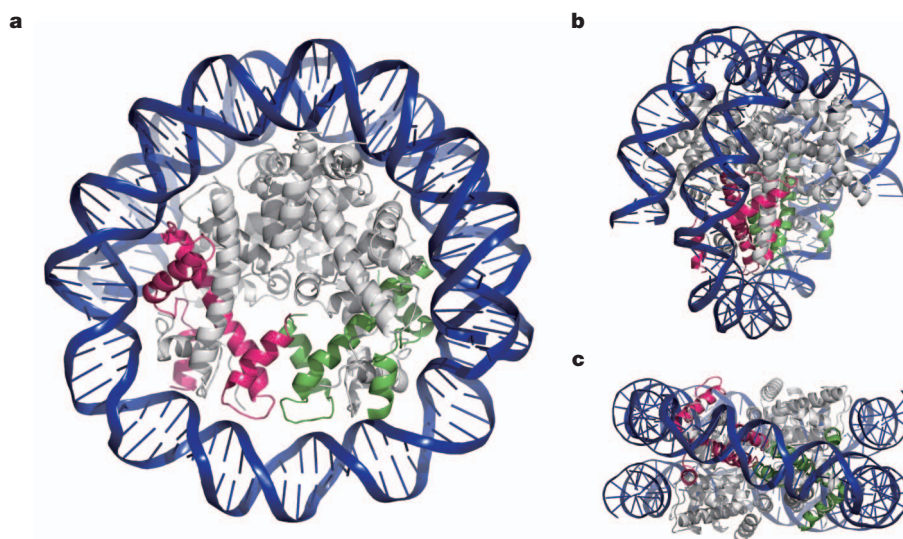
Hiroaki Tachiwana<sup>1\*</sup>, Wataru Kagawa<sup>1\*</sup>, Tatsuya Shiga<sup>1\*</sup>, Akihisa Osakabe<sup>1</sup>, Yuta Miya<sup>1</sup>, Kengo Saito<sup>1</sup>, Yoko Hayashi-Takanaka<sup>2</sup>, Takashi Oda<sup>3</sup>, Mamoru Sato<sup>3</sup>, Sam-Yong Park<sup>4</sup>, Hiroshi Kimura<sup>2</sup> & Hitoshi Kurumizaka<sup>1</sup>

In eukaryotes, accurate chromosome segregation during mitosis and meiosis is coordinated by kinetochores, which are unique chromosomal sites for microtubule attachment<sup>1,2</sup>. Centromeres specify the kinetochore formation sites on individual chromosomes, and are epigenetically marked by the assembly of nucleosomes containing the centromere-specific histone H3 variant, CENP-A<sup>3–12</sup>. Although the underlying mechanism is unclear, centromere inheritance is probably dictated by the architecture of the centromeric nucleosome. Here we report the crystal structure of the human centromeric nucleosome containing CENP-A and its cognate  $\alpha$ -satellite DNA derivative (147 base pairs). In the human CENP-A nucleosome, the DNA is wrapped around the histone octamer, consisting of two each of histones H2A, H2B, H4 and CENP-A, in a left-handed orientation. However, unlike the canonical H3 nucleosome, only the central 121 base pairs of the DNA are visible. The thirteen base pairs from both ends of the DNA are invisible in the crystal structure, and the  $\alpha$ N helix of CENP-A is shorter than that of H3, which is known to be important for the orientation of the DNA ends in the canonical H3 nucleosome<sup>13</sup>. A structural comparison of the CENP-A and H3 nucleosomes revealed that CENP-A contains two extra amino acid residues (Arg 80 and Gly 81) in the loop 1 region, which is completely exposed to the solvent. Mutations of the CENP-A loop 1 residues reduced CENP-A retention at the centromeres in human

cells. Therefore, the CENP-A loop 1 may function in stabilizing the centromeric chromatin containing CENP-A, possibly by providing a binding site for *trans*-acting factors. The structure provides the first atomic-resolution picture of the centromere-specific nucleosome.

Octasome and hemisome models have been proposed for the CENP-A nucleosome architecture<sup>14</sup>. In the octasome model, two each of histones H2A, H2B, H4 and CENP-A form a histone octamer, and about 150 base pairs of DNA are wrapped in a left-handed orientation around the histone octamer, as in the canonical H3 nucleosomes<sup>15–17</sup>. In the hemisome model, however, one each of histones H2A, H2B, H4 and CENP-A<sup>18,19</sup> form a heterotypic tetramer, and the DNA is wrapped in a right-handed orientation around the tetramer<sup>20</sup>.

To reveal the architecture of the CENP-A nucleosome, we prepared the nucleosome using bacterially expressed human histones H2A, H2B, H4 and CENP-A<sup>16,21</sup> (Supplementary Fig. 1) and a 147-base-pair palindromic DNA, which was designed from a human  $\alpha$ -satellite sequence containing a binding site for the centromeric protein, CENP-B (Supplementary Fig. 2). The CENP-A nucleosome was reconstituted by a salt-dialysis method, and crystallized as described (Supplementary Methods). The structure was determined at 3.6 Å resolution (Fig. 1, Supplementary Table 1). The crystal structure revealed a histone octamer containing two each of histones H2A, H2B, H4 and CENP-A, with the DNA wrapped in a left-handed orientation around the histone octamer



**Figure 1 | Crystal structure of the human CENP-A nucleosome.** **a–c**, Three views of the CENP-A nucleosome structure are presented. **a**, View in the axis of the DNA supercoil; **b**, **c**, views from the side of the DNA supercoil. Two CENP-

A molecules are shown in magenta and green, respectively. The central 121-base-pair DNA region, which is visible in the CENP-A nucleosome structure, is shown in dark blue.

<sup>1</sup>Laboratory of Structural Biology, Graduate School of Advanced Science and Engineering, Waseda University, 2-2 Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8480, Japan. <sup>2</sup>Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamada-oka, Suita, Osaka 565-0871, Japan. <sup>3</sup>Division of Macromolecular Crystallography, Department of Supramolecular Biology, Graduate School of Nanobioscience, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan. <sup>4</sup>Protein Design Laboratory, Department of Supramolecular Biology, Graduate School of Nanobioscience, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan.

\*These authors equally contributed to this work.



(Fig. 1). The overall structure is quite similar to those of previously solved nucleosomes containing other histone H3 variants<sup>13,22,23</sup>. The left-handed DNA wrapping in the crystal structure was also observed in biochemical experiments. CENP-A oligonucleosomes assembled on circular plasmid DNA by a salt-dialysis method or in the presence of human histone chaperones, NAP1 and somatic nuclear autoantigenic sperm protein (sNASP)<sup>24,25</sup>, under physiological salt conditions, introduce left-handed (negative) supercoils in the DNA (Supplementary Figs 3 and 4). These results strongly indicate that the DNA wrapped in a left-handed orientation is the predominant form in the human CENP-A nucleosome.

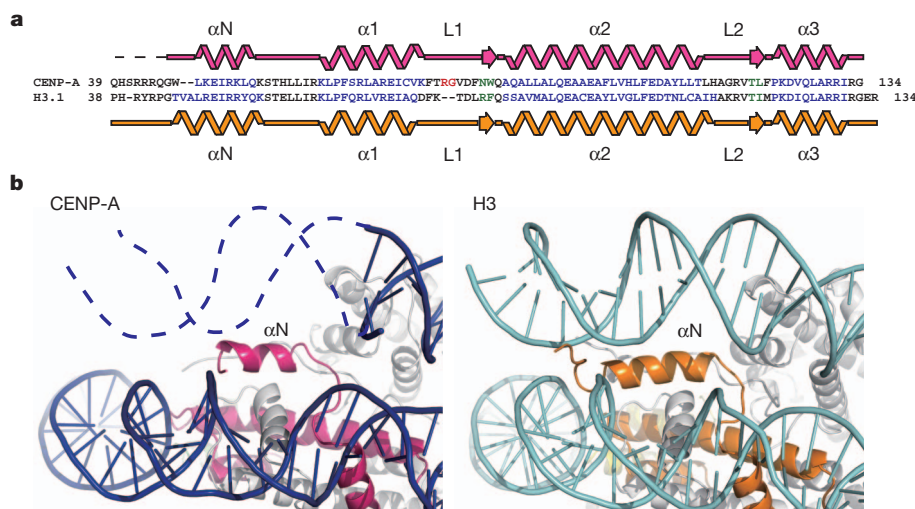
In the free CENP-A–H4 structure reported previously<sup>26</sup>, the CENP-A–CENP-A interface is substantially rotated relative to the H3–H3 interface, indicating that the CENP-A–H4 tetramer may be more compact than the H3–H4 tetramer. However, this specific shape of the CENP-A–CENP-A interface may only be observed in the free CENP-A–H4 tetramer, because the CENP-A–CENP-A interface in the present structure was nearly identical to that of the H3–H3 interface in the canonical H3 nucleosome. Consistently, small-angle X-ray scattering (SAXS) measurements of CENP-A and H3 nucleosomes generated nearly identical SAXS curves and distance distribution functions (Supplementary Fig. 5). These results indicate that the global structures of the CENP-A and H3 nucleosomes are quite similar. In addition, a compact CENP-A nucleosome model containing 147 base-pair DNA constructed from the free H3–H4 tetramer structure<sup>26</sup> generated a significantly different SAXS curve and distance distribution function from those of a CENP-A nucleosome model containing 147 base pair DNA based on the CENP-A nucleosome crystal structure, which was very similar to the experimental data (Supplementary Fig. 6). Therefore, the CENP-A nucleosome structure is probably not compacted as was suggested previously<sup>26</sup>.

Although the overall structure of the CENP-A nucleosome resembles that of the canonical H3 nucleosome, in the CENP-A nucleosome structure, only the central 121 base pairs of the DNA are visible, and thus the thirteen base pairs from both ends of the DNA are disordered (Fig. 1 and Supplementary Fig. 7). This observation indicates that the DNA regions at the entrance and exit of the CENP-A nucleosome lack a fixed conformation. This is consistent with the previous report showing that the DNA segments at the entrance and exit of the CENP-A nucleosome are more flexible than those of the canonical H3 nucleosome<sup>26–28</sup>. We found that the CENP-A nucleosome can be reconstituted on a 121 base pair DNA (Supplementary Fig. 8). The CENP-A nucleosome induced supercoils into plasmid DNA less efficiently than the H3 nucleosome, which

also indicates that the DNA is partially unwrapped in the CENP-A nucleosomes (Supplementary Fig. 9). Furthermore, the exonuclease assay revealed that the DNA ends of the CENP-A nucleosomes were more susceptible to exonuclease III digestion, compared to those of the H3 nucleosome (Supplementary Fig. 10). Similar results were obtained with other DNA sequences without the CENP-B box (Supplementary Fig. 10). Moreover, the SAXS data showed that the maximum dimension  $D_{\max}$  ( $D_{\max} = 165 \text{ \AA}$ ) of the CENP-A nucleosome is slightly longer than that of the H3 nucleosome ( $D_{\max} = 147 \text{ \AA}$ ), which probably reflects the unwrapped conformations of the thirteen base pairs from both ends of the DNA in the CENP-A nucleosome (Supplementary Fig. 5).

This difference in the DNA end structures can be explained by the structural differences between the amino-terminal regions of CENP-A and H3 (Fig. 2a). Previous crystal structures of the canonical H3 nucleosome revealed that the loop segment preceding the  $\alpha$ N helix of H3 interacts with the ends of the DNA, and seems to stabilize their orientations (Fig. 2b)<sup>13,22,23</sup>. Thus, the length of the  $\alpha$ N helix seems to be an important factor for maintaining the DNA orientation at the entrance and exit of nucleosomes. The  $\alpha$ N helix of CENP-A is at least one helical turn shorter than that of H3, and the preceding region is completely disordered (Fig. 2b). The DNA conformations at the entrance and exit of the nucleosome are clearly related to the organization of the nucleosomes<sup>29</sup>, particularly within the heterochromatin, where the nucleosomes are presumably tightly packed. A previous study indicated that the centromeric DNA binding protein, CENP-B, binds efficiently to its recognition sequence<sup>30</sup>, if the sequence is located near the entrance or exit of the CENP-A nucleosome<sup>16</sup>. The flexible nature of both ends of the DNA in the CENP-A nucleosome structure may be an inherent property that facilitates the binding of CENP-B, and possibly other centromeric DNA-binding proteins. Thus, the  $\alpha$ N helix of CENP-A may have a key role in allowing the DNA to adopt the specific conformations at the entrance and exit of the nucleosome, which are not observed in the canonical H3 nucleosome.

It is notable that CENP-A–DNA contacts may exist at the flexible DNA ends of the CENP-A nucleosome. In the exonuclease digestion experiments (Supplementary Fig. 9), we observed DNA fragments longer than 121 base pairs, suggesting that the CENP-A–DNA contacts may extend beyond 121 base pairs of DNA. Furthermore, the CENP-A nucleosome reconstitution efficiency was slightly lower with a 121-base-pair DNA, compared to that with a 147-base-pair DNA (Supplementary Fig. 8). A possible interpretation is that the CENP-A–DNA contacts extending beyond 121 base pairs of DNA may stabilize



**Figure 2 | Structure of the DNA entrance and exit of the human CENP-A nucleosome.** **a**, Secondary structure of CENP-A in the nucleosome. The sequences of human CENP-A and H3 are aligned with the secondary structure elements. **b**, Close-up views of the  $\alpha$ N helices and the DNA edge regions of the

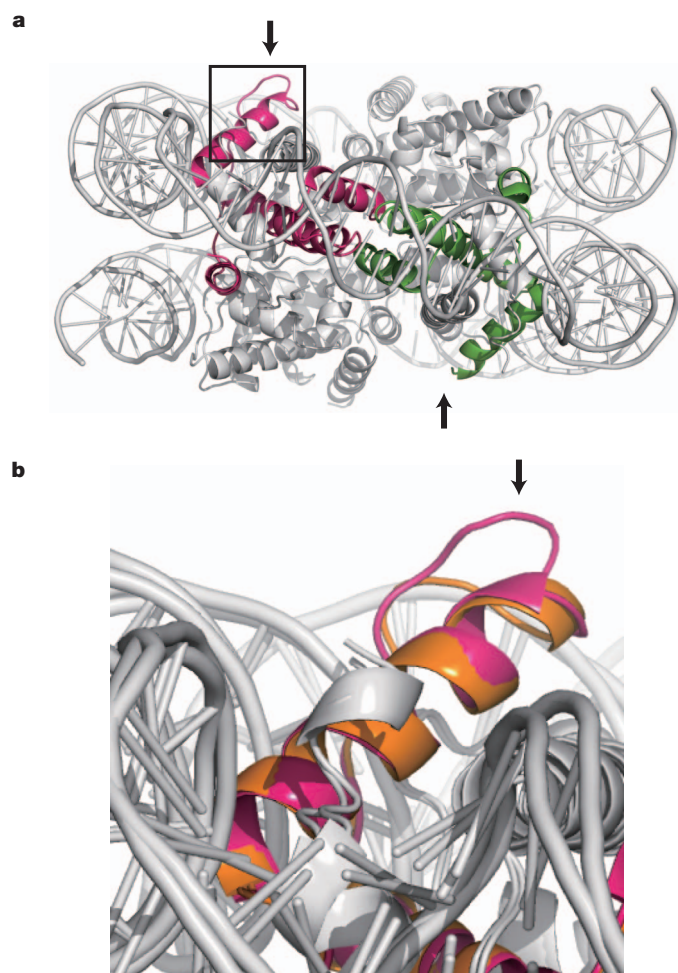
CENP-A (left panel) and H3 (right panel) nucleosomes. The dashed line in the left panel shows the DNA region that is not visible in the crystal structure. The CENP-A and H3 molecules are shown in magenta and orange, respectively.

the CENP-A nucleosome, resulting in higher reconstitution efficiencies on a 147-base-pair DNA.

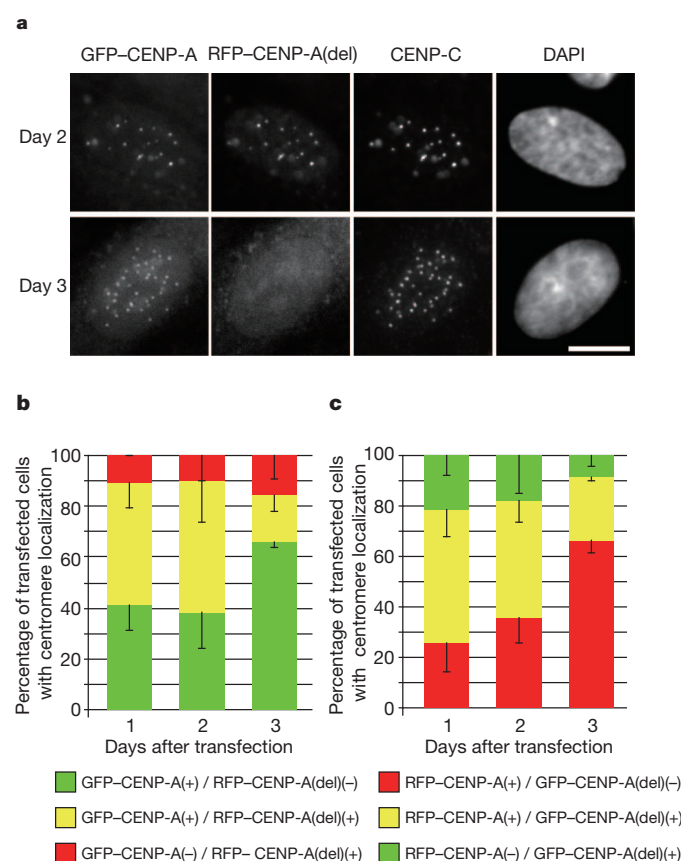
A superimposition of CENP-A and H3 reveals a clear difference in the loop 1 region (residues Phe 78–Phe 84 of CENP-A), where CENP-A has two extra amino acid residues (Arg 80 and Gly 81) compared to H3 (Fig. 2a). The CENP-A loop 1 protrudes from the CENP-A nucleosome, and the Arg 80 and Gly 81 residues are located at the tip of the loop (Fig. 3 and Supplementary Fig. 11). In the free CENP-A–H4 tetramer structure, the loop 1 region is more flexible than the other CENP-A regions, as judged from the B-factors. By contrast, in the CENP-A nucleosome, the B-factors of the loop 1 region are similar to those of the other regions (Supplementary Fig. 12), indicating that CENP-A nucleosome formation may stabilize the loop 1 region. The tip of the loop 1 region is solvent-accessible (Fig. 3), and may function as a binding site for *trans*-acting factors that interact directly with the CENP-A nucleosome. To test the functional significance of the CENP-A Arg 80 and Gly 81 residues, we co-expressed the green fluorescent protein (GFP)-tagged CENP-A and the red fluorescent protein (RFP)-tagged CENP-A(del), in which the Arg 80 and Gly 81 residues were deleted, in human-telomerase-immortalized retina pigment epithelial (hTERT-RPE1) cells. Within 1 or 2 days after transfection, both GFP-tagged CENP-A and RFP-tagged CENP-A(del) were recruited to the centromeres, which were identified by a constitutive centromere protein, CENP-C (Fig. 4a). This result indicates that the Arg 80 and Gly 81

residues are not essential for targeting CENP-A to the centromeres. However, 3 days after transfection, the number of cells in which the CENP-A(del) signals were detected at the centromeres was markedly reduced (Fig. 4b). Concomitantly, the number of cells with the CENP-A signal alone increased (Fig. 4b). Similar results were obtained when the fluorescent labels were swapped between CENP-A and CENP-A(del), showing that the phenomenon does not depend on the fusion partner (Fig. 4c). These results indicate that CENP-A(del) is less stably incorporated into centromeres, compared to CENP-A. In addition, two CENP-A mutants, one containing the Arg 80–Gly 81 to Ala 80–Ala 81 substitution (CENP-A(A80A81)) and another with the Val 82–Asp 83 deletion, which disrupts the Arg 80–Gly 81 protrusion (CENP-A(del82–83)), were targeted to centromeres at levels comparable to those of CENP-A, 1 day after transfection. The number of cells retaining the CENP-A mutants at the centromeres also decreased, 3 days after transfection (Supplementary Fig. 13a, b). However, like CENP-A, the CENP-A mutant containing the Val 82–Asp 83 to Ala 82–Ala 83 substitution (CENP-A(A82A83)) remained at the centromeres, 3 days after transfection (Supplementary Fig. 13c). Thus, the Arg 80 and Gly 81 residues and the size of the protruding loop 1 are critical for stable CENP-A retention at centromeres.

There has been much debate over the CENP-A nucleosome structure and its role in the centromere-specific chromatin structure. Because CENP-A has lower sequence homology to H3, compared to



**Figure 3 | Structural differences in the loop 1 regions between CENP-A and H3.** **a**, Side view of the CENP-A nucleosome. The CENP-A molecules are shown in magenta and green. The box indicates the region enlarged in panel **b**. **b**, Superimposition of the CENP-A (magenta) and H3 (orange) loop 1 regions. Arrows indicate the tip of the CENP-A loop 1 containing the Arg 80 and Gly 81 residues.



**Figure 4 | Less stable association of CENP-A(del) with the centromere.**

**a**, Fluorescence images. hTERT-RPE1 cells were transfected with GFP-tagged CENP-A and RFP-tagged CENP-A(del), fixed, and stained with anti-CENP-C (Cy5) and DAPI. Bar, 10  $\mu$ m. **b**, Quantitative data. Using images such as those in panel **a**, the numbers of transfected hTERT-RPE1 cells showing GFP-CENP-A, RFP-CENP-A(del), or both at centromeres were counted ( $n > 28$ ), and the average percentages from three independent transfections were plotted with the standard deviations. **c**, hTERT-RPE1 cells were transfected with GFP-tagged CENP-A(del) and RFP-tagged CENP-A, and were analysed as described in panel **b** ( $n > 20$ ).



other H3 variants, the possibility of a CENP-A nucleosome composed of one of each core histone (hemisome) is an attractive proposition. However, our findings support the octasome model for the CENP-A nucleosome. It is still possible that both types of CENP-A nucleosomes, octasome and hemisome, coexist in the functional centromeric chromatin *in vivo*. We also cannot exclude the possibility that CENP-A hemisomes can be reconstituted under different conditions and/or with factor(s) required for their assembly. Nevertheless, the present structure suggests that the fundamental principles involved in nucleosome formation are likely to be similar among the H3 variants, including CENP-A. The flexibility exclusively observed in the DNA regions located at the entrance and the exit of the CENP-A nucleosome and the loop 1 region protruding from the CENP-A nucleosome may have an essential role in the centromeric chromatin architecture.

## METHODS SUMMARY

Human CENP-A, H2A, H2B, H3.1 and H4 were overexpressed in *Escherichia coli* cells, and were purified by a method described previously<sup>21,23–25</sup>. Details are provided in Methods. The 147-base-pair DNA used in the CENP-A nucleosome reconstitution was prepared by self-ligation with the 71-base-pair fragment of a human  $\alpha$ -satellite sequence<sup>16,21</sup>, containing an extra 5-base overhang, 5'-GTAAC-3', for the cohesive end. The resultant 147-base-pair DNA contained the CENP-B box near both edges, and an A:A mismatch was located at the centre of the DNA (Supplementary Fig. 2). The preparation, crystallization and structural determination of the CENP-A nucleosome are described in Methods. Analyses of the fluorescent protein-tagged CENP-A or CENP-A mutant incorporation at centromeres were performed using hTERT-RPE1 cells. Details are described in Methods.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 29 August 2010; accepted 1 June 2011.

Published online 10 July 2011.

- Cheeseman, I. M. & Desai, A. Molecular architecture of the kinetochore-microtubule interface. *Nature Rev. Mol. Cell Biol.* **9**, 33–46 (2008).
- Santaguida, S. & Musacchio, A. The life and miracles of kinetocores. *EMBO J.* **28**, 2511–2531 (2009).
- Palmer, D. K., O'Day, K., Wener, M. H., Andrews, B. S. & Margolis, R. L. A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *J. Cell Biol.* **104**, 805–815 (1987).
- Stoler, S., Keith, K. C., Curnick, K. E. & Fitzgerald-Hayes, M. A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. *Genes Dev.* **9**, 573–586 (1995).
- Meluh, P. B., Yang, P., Glowczewski, L., Koshland, D. & Smith, M. M. Cse4p is a component of the core centromere of *Saccharomyces cerevisiae*. *Cell* **94**, 607–613 (1998).
- Buchwitz, B. J., Ahmad, K., Moore, L. L., Roth, M. B. & Henikoff, S. A histone-H3-like protein in *C. elegans*. *Nature* **401**, 547–548 (1999).
- Henikoff, S., Ahmad, K., Platero, J. S. & van Steensel, B. Heterochromatic deposition of centromeric histone H3-like proteins. *Proc. Natl Acad. Sci. USA* **97**, 716–721 (2000).
- Howman, E. V. *et al.* Early disruption of centromeric chromatin organization in centromere protein A (*Cenpa*) null mice. *Proc. Natl Acad. Sci. USA* **97**, 1148–1153 (2000).
- Takahashi, K., Chen, E. S. & Yanagida, M. Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. *Science* **288**, 2215–2219 (2000).
- Blower, M. D. & Karpen, G. H. The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nature Cell Biol.* **3**, 730–739 (2001).
- Oegema, K., Desai, A., Rybina, S., Kirkham, M. & Hyman, A. A. Functional analysis of kinetochore assembly in *Caenorhabditis elegans*. *J. Cell Biol.* **153**, 1209–1226 (2001).
- Régnier, V. *et al.* CENP-A is required for accurate chromosome segregation and sustained kinetochore association of BubR1. *Mol. Cell Biol.* **25**, 3967–3981 (2005).
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
- Talbert, P. B. & Henikoff, S. Histone variants—ancient wrap artists of the epigenome. *Nature Rev. Mol. Cell Biol.* **11**, 264–275 (2010).
- Yoda, K. *et al.* Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution *in vitro*. *Proc. Natl Acad. Sci. USA* **97**, 7266–7271 (2000).
- Tanaka, Y. *et al.* Human centromere protein B induces translational positioning of nucleosomes on alpha-satellite sequences. *J. Biol. Chem.* **280**, 41609–41618 (2005).
- Camahort, R. *et al.* Cse4 is part of an octameric nucleosome in budding yeast. *Mol. Cell* **35**, 794–805 (2009).
- Dalal, Y., Wang, H., Lindsay, S. & Henikoff, S. Tetrameric structure of centromeric nucleosomes in interphase *Drosophila* cells. *PLoS Biol.* **5**, e218 (2007).
- Dalal, Y., Furuyama, T., Vermaak, D. & Henikoff, S. Structure, dynamics, and evolution of centromeric nucleosomes. *Proc. Natl Acad. Sci. USA* **104**, 15974–15981 (2007).
- Furuyama, T. & Henikoff, S. Centromeric nucleosomes induce positive DNA supercoils. *Cell* **138**, 104–113 (2009).
- Tanaka, Y. *et al.* Expression and purification of recombinant human histones. *Methods* **33**, 3–11 (2004).
- Tsunaka, Y., Kajimura, N., Tate, S. & Morikawa, K. Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle. *Nucleic Acids Res.* **33**, 3424–3434 (2005).
- Tachiwana, H. *et al.* Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. *Proc. Natl Acad. Sci. USA* **107**, 10454–10459 (2010).
- Tachiwana, H., Osakabe, A., Kimura, H. & Kurumizaka, H. Nucleosome formation with the testis-specific histone H3 variant, H3t, by human nucleosome assembly proteins *in vitro*. *Nucleic Acids Res.* **36**, 2208–2218 (2008).
- Osakabe, A. *et al.* Nucleosome formation activity of human somatic nuclear autoantigenic sperm protein (sNASP). *J. Biol. Chem.* **285**, 11913–11921 (2010).
- Sekulic, N., Bassett, E. A., Rogers, D. J. & Black, B. E. The structure of (CENP-A-H4)<sub>2</sub> reveals physical features that mark centromeres. *Nature* **467**, 347–351 (2010).
- Conde e Silva, N. *et al.* CENP-A-containing nucleosomes: easier disassembly versus exclusive centromeric localization. *J. Mol. Biol.* **370**, 555–573 (2007).
- Kingston, I. J., Yung, J. S. & Singleton, M. R. Biophysical characterisation of the centromere-specific nucleosome from budding yeast. *J. Biol. Chem.* **286**, 4021–4026 (2011).
- Schalch, T., Duda, S., Sargent, D. F. & Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436**, 138–141 (2005).
- Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. & Okazaki, T. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J. Cell Biol.* **109**, 1963–1973 (1989).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the beamline scientists, N. Shimizu, Y. Kawano, M. Makino and T. Hikima, for their assistance with data collection at the BL41XU and BL45XU beamlines of SPring-8. We also thank R. Matsumoto for technical assistance, K. Yoda for anti-CENP-C, and T. Fukagawa and Y. Hiraoka for discussions. This work was supported in part by Grants-in-Aid from the Japanese Society for the Promotion of Science (JSPS), and the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. H.K. was also supported by the Waseda Research Institute for Science and Engineering.

**Author Contributions** H.T., T.S., A.O. and Y.M. purified the histones and CENP-A, crystallized the CENP-A nucleosome, and performed biochemical analyses. H.T., W.K., K.S. and T.S. collected X-ray diffraction data, and H.T., W.K., and S.-Y.P. performed the structural analysis of the CENP-A nucleosome. H.T., A.O., Y.H.-T. and H.Ki. performed the cell biological experiments. T.O., H.T., W.K. and M.S. performed SAXS analysis. H.Ku. conceived, designed and supervised all of the work, and H.Ku., W.K. and H.T. wrote the paper. All of the authors discussed the results and commented on the manuscript.

**Author Information** The atomic coordinates of the CENP-A nucleosome have been deposited in the Protein Data Bank, under the accession code 3AN2. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to H.Ku. ([kurumizaka@waseda.jp](mailto:kurumizaka@waseda.jp)).



## METHODS

**Overexpression of human histones.** Human histones H2A and H2B were produced in *Escherichia coli* BL21(DE3) cells, and histone H4 was produced in *E. coli* JM109(DE3) cells. Human CENP-A was produced in *E. coli* DH5 $\alpha$  cells. All histones and CENP-A were produced in *E. coli* cells in the absence of T7 RNA polymerase by omitting the addition of isopropyl- $\beta$ -D-thiogalactopyranoside, which induces the T7 RNA polymerase production in BL21(DE3) and JM109(DE3) cells. All histones and CENP-A were produced as N-terminal His<sub>6</sub>-tagged proteins, as described previously<sup>21</sup>. The His<sub>6</sub> tags of all histones were removed by thrombin protease digestion, leaving a Gly-Ser-His sequence at the N-terminal end of each histone.

For the purpose of structural determination, selenomethionine (Se-Met)-substituted H2B was produced in *E. coli* B834(DE3) cells, using the pET15b vector system (Novagen). The B834(DE3) cells were grown in 100 ml of LB medium for 4 h at 37 °C. The cells were collected and transferred into 300 ml of M9 medium (+50  $\mu$ g ml<sup>-1</sup> Se-Met). After 12 h growth at 37 °C, the 300-ml culture was added to 21 of M9 medium (+50  $\mu$ g ml<sup>-1</sup> Se-Met), and the culture was continued at 37 °C. When the cell density reached 0.5 (*D*<sub>600</sub>), isopropyl- $\beta$ -D-thiogalactopyranoside (final concentration 1 mM) was added, to induce the expression of H2B. The cells were grown further at 37 °C for 12 h.

**Purification of human histones.** The cells producing recombinant histones were collected, and were resuspended in 50 ml of buffer A (50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 1 mM PMSF and 5% glycerol). The cells were disrupted by two rounds of sonication for 200 s each. The cell lysates were centrifuged at 27,216g for 20 min at 4 °C. The supernatants were discarded, and the pellet containing the His<sub>6</sub>-tagged histones was resuspended in 50 ml of buffer A, containing 7 M guanidine hydrochloride. The samples were rotated for 12 h at 4 °C, and the supernatants were recovered by centrifugation at 27,216g for 20 min at 4 °C. The supernatants containing the His<sub>6</sub>-tagged histones were combined with 4 ml (50% slurry) of nickel-nitrilotriacetic acid (Ni-NTA) agarose resin (Qiagen), and the samples were rotated for 1 h at 4 °C. The agarose beads were then washed with 100 ml of buffer B (50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 6 M urea, 5 mM imidazole, and 5% glycerol). The His<sub>6</sub>-tagged histones were eluted by a 100-ml linear gradient of 5 to 500 mM imidazole in buffer B, and the samples were dialysed against buffer C (5 mM Tris-HCl (pH 7.5) and 2 mM 2-mercaptoethanol). The N-terminal His<sub>6</sub> tags were removed from the histones by thrombin protease treatment (1 unit mg<sup>-1</sup> of histones; GE Healthcare) at room temperature for 3 h. The removal of the His<sub>6</sub> tags was confirmed by SDS-16% polyacrylamide gel electrophoresis (PAGE); the recombinant histones without the His<sub>6</sub> tag migrated faster than the His<sub>6</sub>-tagged histones. After the His<sub>6</sub> tag was uncoupled, each histone was subjected to Mono S column chromatography (GE Healthcare). The column was washed with buffer D (20 mM sodium acetate (pH 5.2), 200 mM NaCl, 5 mM 2-mercaptoethanol, 1 mM EDTA, and 6 M urea), and each histone was eluted by a linear gradient of 200 to 800 mM NaCl in buffer D. The purified histones were dialysed against water, and were freeze-dried.

**Preparation of DNAs.** The 147-base-pair DNA, which was used for reconstituting the CENP-A nucleosome, is a derivative of the human  $\alpha$ -satellite DNA (sat4)<sup>21</sup>. The EcoRI site (GAATTC) of the sat4 sequence was replaced by a BstPI site (GGTAACC). The 71-mer DNA fragment containing the 5' half of the sat4 sequence, with the CENP-B box at the edge, was ligated in tandem in the plasmid (p5'Sat4-24). The 71-mer DNA fragment containing an extra 5-base overhang, 5'-GTAAC-3', was prepared for self-ligation according to the method described previously<sup>31</sup>. The 71-mer DNA fragment containing the 5-base overhang was self-ligated, and the palindromic 147-base-pair  $\alpha$ -satellite DNA derivative was prepared. The 147-base-pair DNA sequence is: 5'-ATCCTTCGTTGGAAACG GGATTTCCTTCATTCATGCTAGACAGAAGAATTCCTCAGTAACCTCTTTG TGCTGGTAACACAGCACAAGAAGTTACTGAGAATTCCTCTGCTAGCAT GAAATGAAGAAATCCCGTTTCCAACGAAGGAT-3'.

In this palindromic 147-base-pair  $\alpha$ -satellite DNA derivative, an A:A mismatch was introduced at the centre of the 147-base-pair DNA fragment (underlined).

**Preparation of the CENP-A nucleosome.** The purified H2A-H2B (Se-Met)-CENP-A-H4 (0.9 mg) and the 147-base-pair DNA (1 mg) were mixed in a solution containing 2 M KCl, and the sample was dialysed against dialysis buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 mM dithiothreitol and 2 M KCl). After dialysis at 4 °C for 3 h, the KCl concentration of the dialysis buffer was gradually decreased to 250 mM with a peristaltic pump (0.8 ml min<sup>-1</sup> flow rate). The sample was then dialysed against 10 mM Tris-HCl buffer (pH 7.5), containing 1 mM EDTA, 1 mM dithiothreitol and 250 mM KCl, at 4 °C for 3 h. After this dialysis step, the sample was incubated at 55 °C for 2 h. The CENP-A nucleosome was purified from the free DNA and histones by non-denaturing polyacrylamide gel electrophoresis, using a Prep Cell apparatus (Bio-Rad). The purified CENP-A nucleosome was concentrated, and was dialysed against 20 mM potassium cacodylate buffer (pH 6.0) containing 1 mM EDTA.

**Crystallization and structure determination.** Crystals of the purified CENP-A nucleosome were obtained by the hanging drop method, after mixing equal volumes of the CENP-A nucleosome solution and 20 mM potassium cacodylate buffer (pH 6.0), containing 60–96 mM KCl and 135–144 mM MnCl<sub>2</sub>. The CENP-A nucleosome sample was equilibrated against a reservoir solution of 20 mM potassium cacodylate (pH 6.0), 38–56 mM KCl, and 70–75 mM MnCl<sub>2</sub>. Crystals of the CENP-A nucleosome were soaked in a cryoprotectant solution, containing 20 mM potassium cacodylate (pH 6.0), 47 mM KCl, 72 mM MnCl<sub>2</sub>, 30% polyethylene glycol 400, and 5% trehalose. The crystals were flash-cooled in a stream of N<sub>2</sub> gas (100 K). The CENP-A nucleosome crystals belonged to the monoclinic space group P2<sub>1</sub>, with unit cell constants of *a* = 65.8 Å, *b* = 83.3 Å, *c* = 176.8 Å and  $\beta$  = 100.7°, and contained one nucleosome in the asymmetric unit. High-resolution diffraction data were obtained using the synchrotron radiation source at the beamline BL41XU station of SPring-8, Harima, Japan.

Diffraction data of the CENP-A nucleosome were integrated and scaled with the HKL2000 program<sup>32</sup>. The data were processed with the CCP4 program suite<sup>33</sup>. The structure was solved by the molecular replacement method, using the MOLREP program<sup>34</sup> and the human nucleosome structure (PDB accession number 3AFA) as a search model<sup>23</sup>. Most of the amino acid side chains were clearly visible in the map initially calculated at 3.6 Å resolution. Rigid body refinement of the obtained solution was performed using the CNS program<sup>35</sup>. Further structural refinement consisted of iterative rounds of energy minimization and B factor refinement using the CNS program<sup>35</sup>, and model building using the COOT program<sup>36</sup>. The Ramachandran plot of the final structure showed 98.7% of the residues in the most favourable and additional allowed regions, and no residues in the disallowed region. Summaries of the data collection and refinement statistics are provided in Supplementary Table 1. All structure figures were created using the PyMOL program (<http://pymol.org>). The atomic coordinates of the CENP-A nucleosome have been deposited in the Protein Data Bank, with the ID code 3AN2.

**Supercoiling assay.** Salt-dialysis supercoiling assay. Relaxed plasmid DNA (500 ng) was mixed with 0, 125, 250, 500 and 1,000 ng of histone octamer in 5  $\mu$ l of 20 mM Tris-HCl (pH 7.5) buffer, containing 1 mM EDTA, 0.2 mg ml<sup>-1</sup> BSA, and 2 M NaCl. The samples were then incubated at 37 °C for 30 min. The NaCl concentration of the sample was reduced to 1 M, 0.8 M, 0.67 M, and 0.2 M by adding dilution buffer, containing 20 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.2 mg ml<sup>-1</sup> BSA, 5 mM MgCl<sub>2</sub>, and 0.06 U  $\mu$ l<sup>-1</sup> calf (Invitrogen) or wheat germ (Promega) topoisomerase I. The samples were incubated at 37 °C for 30 min in each dilution step.

Chaperone-mediated supercoiling assay. NAP1 or sNAP (0.25, 0.5, and 1.0  $\mu$ M) was pre-incubated with H2A-H2B (150 ng) and CENP-A-H4 (150 ng) at 37 °C for 15 min. Supercoiled plasmid DNA (100 ng), which was relaxed with a topoisomerase I solution (10 mM Tris-HCl (pH 8.0), 2 mM MgCl<sub>2</sub>, 5 mM dithiothreitol, and 2 U  $\mu$ l<sup>-1</sup> wheat germ topoisomerase I (Promega)), was added to the reaction mixture. The samples were then incubated at 37 °C for 60 min in 10 mM Tris-HCl (pH 8.0) buffer, containing 140 mM NaCl, 2 mM MgCl<sub>2</sub>, and 5 mM dithiothreitol, followed by an incubation at 42 °C for 60 min.

In both the salt-dialysis and chaperone-mediated assays, after the reaction, the samples were treated with 50  $\mu$ l of a proteinase K solution (20 mM Tris-HCl (pH 8.0), 20 mM EDTA, 0.5% SDS, and 0.5 mg ml<sup>-1</sup> proteinase K (Roche)) at 37 °C for 30 min. The DNA was extracted with phenol/chloroform. The DNA was then precipitated by ethanol, and was analysed by one-dimensional gel electrophoresis on a 1% agarose gel in 1  $\times$  TAE buffer (for the salt-dialysis assay, 1.3 V cm<sup>-1</sup> for 15.5 h) or 1  $\times$  TBE buffer (for the chaperone-mediated assay, 1.3 V cm<sup>-1</sup> for 15.5 h). For the two-dimensional gel electrophoresis, the DNA was electrophoresed on a 0.7% agarose gel in 1  $\times$  TBE buffer (for the salt-dialysis assay, 2 V cm<sup>-1</sup> for 7 h) or a 1% agarose gel in 1  $\times$  TBE buffer (for the chaperone-mediated assay, 1.3 V cm<sup>-1</sup> for 15 h) for the first dimension. The gel was then soaked in 1  $\times$  TBE buffer containing 4 mg l<sup>-1</sup> of chloroquine for 3 h. The samples were subsequently electrophoresed in 1  $\times$  TBE buffer containing 4 mg l<sup>-1</sup> of chloroquine (1.3 V cm<sup>-1</sup> for 12 h (for the salt-dialysis assay) or 1.3 V cm<sup>-1</sup> for 15 h (for the chaperone-mediated assay)) for the second dimension. The DNA was visualized by SYBR Gold (Invitrogen) staining.

**Nucleosome reconstitution by the salt-dialysis method for biochemical analyses.** The purified H2A-H2B-CENP-A-H4 or H2A-H2B-H3-H4 octamer was mixed with a DNA fragment (300  $\mu$ g, 121-base-pair DNA or 147-base-pair DNA) in a solution containing 2 M KCl (376  $\mu$ l). The amounts of histone octamers were 420  $\mu$ g for the 121-base-pair DNA and 384  $\mu$ g for the 147-base-pair DNA. Nucleosomes were reconstituted and prepared by the same method as described in the 'Preparation of the CENP-A nucleosome' section.

**Competitive nucleosome assembly assay.** The purified H2A-H2B-CENP-A-H4 octamer (14, 28, 42 or 56  $\mu$ g) was incubated in the presence of both the 147-base-pair DNA (24  $\mu$ g) and 121-base-pair DNA (20  $\mu$ g), in a solution containing 2 M KCl, and the sample was dialysed against dialysis buffer (10 mM Tris-HCl

(pH 7.5), 1 mM EDTA, 1 mM dithiothreitol, and 2 M KCl). After dialysis at 4 °C for 3 h, the KCl concentration of the dialysis buffer was gradually decreased to 250 mM with a peristaltic pump (0.8 ml min<sup>-1</sup> flow rate). The sample was then dialysed against 10 mM Tris-HCl buffer (pH 7.5), containing 1 mM EDTA, 1 mM dithiothreitol, and 250 mM KCl, at 4 °C for 3 h. The CENP-A nucleosomes were then analysed by 6% PAGE in 0.2× TBE buffer (18 mM Tris base, 18 mM boric acid, and 0.4 mM EDTA) at 16 V cm<sup>-1</sup> for 1 h, followed by ethidium bromide staining.

**Nucleosome disruption assay.** The CENP-A nucleosomes were reconstituted with a 121-base-pair or 147-base-pair palindromic  $\alpha$ -satellite derivative, by the salt dialysis method. The 121-base-pair DNA lacks the 13-base-pair regions from both edges of the 147-base-pair DNA used in the crystallography of the CENP-A nucleosome. The rest of the 121-base-pair DNA sequence is identical to the 147-base-pair palindromic  $\alpha$ -satellite derivative. The CENP-A nucleosomes (150 ng) were incubated at 37 °C, 57 °C, 67 °C, 70 °C, or 73 °C for 15 min in the presence of supercoiled plasmid DNA (100 ng). After the incubation, the CENP-A nucleosomes that were not disrupted were separated by non-denaturing 6% PAGE, and were visualized by ethidium bromide staining. The relative band intensities for the CENP-A nucleosomes were quantified and plotted against the temperature.

**Exonuclease assay.** The reconstituted CENP-A or H3 nucleosomes were treated with 3 units of *Escherichia coli* exonuclease III (Takara), in 10  $\mu$ l of 50 mM Tris-HCl (pH 8.0), 5 mM MgCl<sub>2</sub>, and 1 mM DTT. After an incubation for 0, 2, 4, or 8 min at 37 °C, the reaction was stopped by the addition of 55  $\mu$ l of proteinase K solution (20 mM Tris-HCl (pH 8.0), 20 mM EDTA, 0.5% SDS, and 0.5 mg ml<sup>-1</sup> proteinase K (Roche)). After a 15 min incubation at room temperature, the DNA was extracted with phenol/chloroform, precipitated with ethanol, dissolved in Hi-Di Formamide (Applied Biosystems), and then analysed by 10% denaturing PAGE with a gel containing 7 M urea in 0.5× TBE buffer (21 V cm<sup>-1</sup> for 1.5 h).

**Small-angle X-ray scattering (SAXS).** SAXS measurements of the reconstituted CENP-A and H3 nucleosomes, in 20 mM Tris-HCl buffer (pH 7.5) containing 1 mM EDTA and 1 mM DTT, were performed at the RIKEN structural biology beam-line I (BL45XU) of SPring-8 (Hyogo, Japan)<sup>37</sup>. Scattering intensities of the nucleosome solutions were measured with an R-Axis IV<sup>++</sup> imaging plate detector at 20 °C with a sample-to-detector distance of 3,529 mm, which was calibrated by the powder diffraction from silver dicosanoate. Circular averaging of the scattering intensities was then performed to obtain the one-dimensional scattering data  $I(q)$  as a function of  $q$  ( $q = 4\pi\sin\theta/\lambda$ ), where  $2\theta$  is the scattering angle and the X-ray wavelength  $\lambda = 0.9$  Å). Three successive measurements were made for each solution, with an exposure time of 60 s. The resultant three data sets were combined after inspections for X-ray radiation damage to the solution and the existence of instrumental artefacts. SAXS measurements of the buffer solution for background subtraction were performed after each measurement of the nucleosome solutions, using the same conditions and procedure as those of the nucleosome solutions. To correct the inter-particle interference effect,  $I(q)$  data were collected at four protein concentrations (0.5, 0.7, 1.0 and 1.3 mg ml<sup>-1</sup>), and extrapolated to zero concentration. The data were processed and analysed using the software applications embedded in the ATSAS package (<http://www.>

[embl-hamburg.de/biosaxs/software.html](http://embl-hamburg.de/biosaxs/software.html)). The radius of gyration,  $R_g$ , was estimated by fitting the  $I(q)$  data using the Guinier approximation  $I(q) = I(0) \exp(-q^2 R_g^2/3)$ , where  $I(0)$  is the forward scattering at the zero scattering angle, in a smaller angle region of  $qR_g < 1.3$ . Error of  $R_g$  was estimated from the least-squares fitting. The distance distribution function  $P(r)$  and its error were calculated by the program GNOM<sup>38</sup>. The maximum dimension  $D_{max}$  was estimated from the  $P(r)$  function as the distance  $r$ , where  $P(r) = 0$  (ref. 39), and its error was estimated from the errors of the  $P(r)$  values around  $P(r) = 0$ .

**Centromere localization of CENP-A and CENP-A mutants.** hTERT-RPE1 cells were transfected with combinations of wild-type CENP-A and CENP-A( $\Delta$ del), in which two amino acid residues (the Arg 80 and Gly 81 residues of the CENP-A loop 1) were deleted, and tagged with either GFP or RFP, using GeneJuice (Merck) according to the manufacturer's instructions. hTERT-RPE1 cells were also transfected with combinations of wild-type CENP-A tagged with RFP and CENP-A( $\Delta$ del82-83) (where the Val 82 and Asp 83 residues of the CENP-A loop 1 were deleted), or CENP-A(A80A81) (where the Arg 80 and Gly 81 residues were replaced by Ala 80 and Ala 81), or CENP-A(A82A83) (where the Val 82 and Asp 83 residues were replaced by Ala 82 and Ala 83), tagged with GFP. The cells were fixed with 4% paraformaldehyde 1–3 days after transfection, permeabilized, and stained with guinea pig anti-CENP-C<sup>40</sup> and donkey Cy5-conjugated anti-guinea pig Ig (Jackson ImmunoResearch). DNA was counterstained with 12.5 ng ml<sup>-1</sup> DAPI. The fluorescence images were collected using an inverted microscope (Ti-E; Nikon) with a  $\times 100$  PlanApo VC numerical aperture (NA) = 1.4 oil-immersion objective lens, or a  $\times 40$  PlanApo NA = 0.95 dry lens, equipped with an EM-CCD camera (iXon+; Andor). The numbers of transfected cells exhibiting the GFP- or RFP-tagged protein, or both, at the centromeres were counted, and the average percentages from three independent transfections were plotted with the standard deviations.

1. Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* **375**, 23–44 (2003).
2. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
3. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
4. Vagin, A. & Teplyakov, A. *MOLREP*: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025 (1997).
5. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
6. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr. D* **66**, 486–501 (2010).
7. Fujisawa, T. *et al.* Small-angle X-ray scattering station at the SPring-8 RIKEN beamline. *J. Appl. Cryst.* **33**, 797–800 (2000).
8. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* **25**, 495–503 (1992).
9. Glatter, O. & Kratky, O. *Small-angle X-ray Scattering* (Academic Press, 1982).
10. Ando, S. *et al.* CENP-A, -B, and -C chromatin complex that contains the I-type  $\alpha$ -satellite array constitutes the prekinetochore in HeLa cells. *Mol. Cell. Biol.* **22**, 2229–2241 (2002).

# Inkjet printing of single-crystal films

Hiromi Minemawari<sup>1</sup>, Toshikazu Yamada<sup>1</sup>, Hiroyuki Matsui<sup>1</sup>, Jun'ya Tsutsumi<sup>1</sup>, Simon Haas<sup>1</sup>, Ryosuke Chiba<sup>1,2</sup>, Reiji Kumai<sup>1,3</sup> & Tatsuo Hasegawa<sup>1</sup>

The use of single crystals has been fundamental to the development of semiconductor microelectronics and solid-state science<sup>1</sup>. Whether based on inorganic<sup>2–5</sup> or organic<sup>6–8</sup> materials, the devices that show the highest performance rely on single-crystal interfaces, with their nearly perfect translational symmetry and exceptionally high chemical purity. Attention has recently been focused on developing simple ways of producing electronic devices by means of printing technologies. 'Printed electronics' is being explored for the manufacture of large-area and flexible electronic devices by the patterned application of functional inks containing soluble or dispersed semiconducting materials<sup>9–11</sup>. However, because of the strong self-organizing tendency of the deposited materials<sup>12,13</sup>, the production of semiconducting thin films of high crystallinity (indispensable for realizing high carrier mobility) may be incompatible with conventional printing processes. Here we develop a method that combines the technique of antisolvent crystallization<sup>14</sup> with inkjet printing to produce organic semiconducting thin films of high crystallinity. Specifically, we show that mixing fine droplets of an antisolvent and a solution of an active semiconducting component within a confined area on an amorphous substrate can trigger the controlled formation of exceptionally uniform single-crystal or polycrystalline thin films that grow at the liquid–air interfaces. Using this approach, we have printed single crystals of the organic semiconductor 2,7-diocetyl[1]benzothieno[3,2-*b*][1]benzothiophene (C<sub>8</sub>-BTBT) (ref. 15), yielding thin-film transistors with average carrier mobilities as high as  $16.4\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$ . This printing technique constitutes a major step towards the use of high-performance single-crystal semiconductor devices for large-area and flexible electronics applications.

Antisolvent crystallization is recognized as the best method of achieving controlled and scalable solidification, which is useful in pharmaceutical science<sup>14</sup>, for example. To achieve this, an 'antisolvent' (a liquid in which a substance is insoluble) is added to the solution of the substance in a solvent that is miscible with the antisolvent. Here we make use of this concept in microliquid inkjet printing processes.

A solution of a semiconductor and an antisolvent for the semiconductor are used as the two kinds of ink; the inks are individually printed at arbitrary positions to form a microliquid intermixture between the inks on the top of substrates. We found that optimized printing conditions enable controlled formation of patterned single-crystal thin films having molecularly flat surfaces, in contrast to conventional inkjet printing processes that produce films with a non-uniform thickness distribution. This is a conceptual extension of the 'double-shot' inkjet printing process that was developed to produce films of charge-transfer compounds that are hardly soluble<sup>16,17</sup>. We used 1,2-dichlorobenzene (DCB) as the solvent and *N,N*-dimethylformamide (DMF) as the antisolvent for the semiconductor C<sub>8</sub>-BTBT. These organic liquids show very different solubilities for C<sub>8</sub>-BTBT (the solubility at 20 °C is 400 times higher in DCB than in DMF), but have similar boiling points and are miscible with one another.

A schematic representation of this printing process is shown in Fig. 1a. We used silicon wafers with 100-nm-thick silicon dioxide layers

as substrates. We produced the wetting/non-wetting surface patterning on the silicon dioxide layers by using a combination of ultraviolet/ozone treatment, hexamethyldisilazane treatment, and photoresist patterning<sup>18</sup>. We used a piezoelectric inkjet printing apparatus with double inkjet printing heads, from which a droplet of 60 picolitres is ejected at a repetition frequency of 500 Hz. In the process, the antisolvent ink (pure anhydrous DMF) is printed first and then overprinted with the solution ink (a 28 mM solution of C<sub>8</sub>-BTBT in DCB). In the formation of all the pieces of film shown in Fig. 1b, 42 shots of antisolvent ink were printed first and then 6 shots of solution ink were overprinted, all within a second. The deposited droplets are confined and intermixed in a predefined hydrophilic area on the upper surface of the substrate.

During the initial stages of film formation, tiny floating bodies begin to form at the surface of the liquid and can be seen in microscope images (Supplementary Movie). Each floating body acts as a nucleus for further crystallization and undergoes subsequent growth to form a larger floating body. These bodies eventually cover the entire surface of the droplet (step 3 in Fig. 1a). A few creases can be seen on the surfaces of the droplets during liquid evaporation, indicating the solid nature of the films (step 4 and Supplementary Fig. 1)<sup>19</sup>.

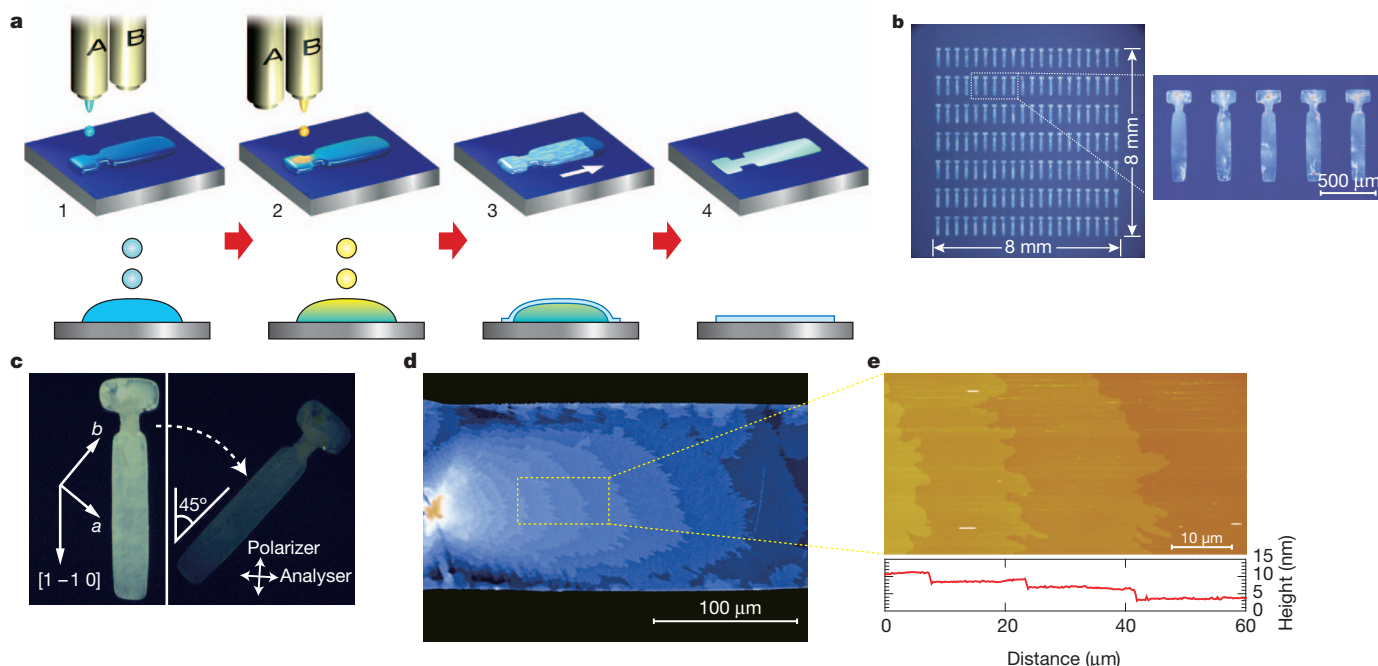
Although nuclei are generated randomly, mostly at the perimeters of the deposited droplet (solid–liquid–air interfaces), we found that nucleation can be controlled through appropriate design of the droplet configuration, which is shaped by the predefined hydrophilic area as well as by the ink volume. For example, a hydrophilic area containing a protuberance, as presented in Fig. 1b, was quite effective in causing local seeding of floating bodies in the protrusive area. We propose that local seeding is associated with the comparatively higher rate of solvent evaporation in areas with a high surface area-to-volume ratio. After seeding, the growing front moves slowly to the other end of the droplet until the large single-domain floating body covers the entire liquid–air interface (see Supplementary Movie).

The solvent then evaporates very slowly, taking about 10–50 times longer than is the case without the solute, most probably because the droplet is completely covered by the solid film. During this slow evaporation, the creases in the films become smoothed out, and films with thickness of about 30–200 nm are eventually obtained on the amorphous substrate. The film adheres tightly to the substrate. The morphology of the films as well as their single-domain nature depends on a variety of printing conditions, such as substrate temperature, the concentration and volume of the solution, the solution–antisolvent ratio and the shape of the hydrophilic area on which the droplet is deposited.

The thickness profile of the film differs markedly from that of conventional inkjet printing deposits. Conventional inkjet printing is known to produce a characteristic thickness distribution in which both ends of the deposit are considerably thicker than its centre, known as the 'coffee-ring effect' (see Supplementary Fig. 2)<sup>20</sup>. The uniform nature of the deposits produced by our process can be ascribed to temporal discrimination between solute crystallization and solvent evaporation within the deposited droplet (see Supplementary Fig. 1)<sup>16</sup>. The occurrence of supersaturation in the intermixed microliquid droplet results

<sup>1</sup>National Institute of Advanced Industrial Science and Technology (AIST), AIST Tsukuba Central 4, Tsukuba 305-8562, Japan. <sup>2</sup>Department of Applied Physics, The University of Tokyo, Hongo 113-8656, Japan. <sup>3</sup>CMRC, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), Tsukuba 305-0801, Japan.





**Figure 1 | Inkjet printing of organic single-crystal thin films.** **a**, Schematic of the process. Antisolvent ink (A) is first inkjet-printed (step 1), and then solution ink (B) is overprinted sequentially to form intermixed droplets confined to a predefined area (step 2). Semiconducting thin films grow at liquid–air interfaces of the droplet (step 3), before the solvent fully evaporates (step 4).

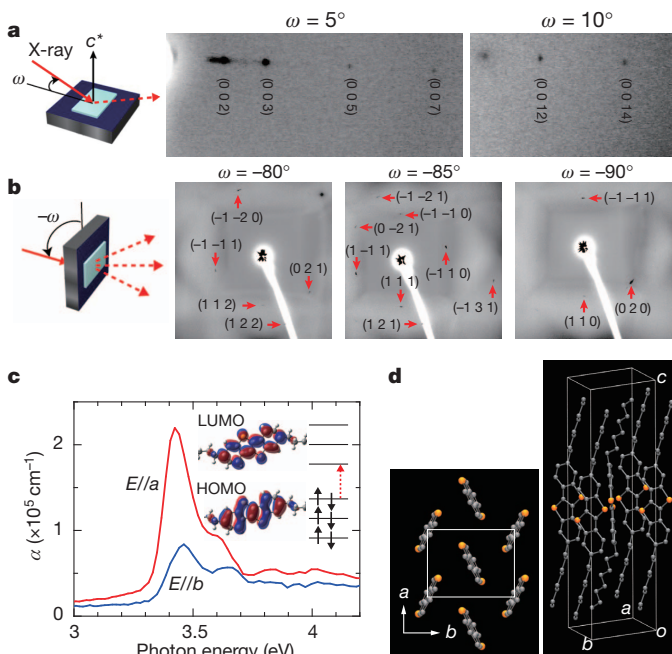
in solute crystallization before solvent evaporation. In the microscope images of the films shown in Fig. 1d, we can see stripe-like features with intervals of several micrometres to several tens of micrometres. Atomic-force microscopy showed that the stripes are associated with the height of the molecular step, which is estimated to be about 2.6–2.8 nm (Fig. 1e)<sup>21</sup>. This value is consistent with the thickness of one molecular layer of C<sub>8</sub>-BTBT ( $c\sin\beta = 2.92$  nm, where  $c$  and  $\beta$  are the unit cell parameters) (ref. 22). We conclude that the stripe-like features are associated with the step-and-terrace structure of C<sub>8</sub>-BTBT.

In images recorded through crossed Nicol prisms, the colour of almost the entire film changes from bright to dark, simultaneously, on rotating the film about an axis perpendicular to the substrate (Fig. 1c). In addition, when we use hydrophilic areas with different configurations such as a simple square, rectangle or circle, we obtained polycrystalline films composed of some crystal domains (see Supplementary Fig. 3). From these observations, we conclude that with appropriate design of both the droplet shape and printing conditions, single-domain crystal films that cover nearly the whole region of the printed deposits could be produced with high probability (Supplementary Fig. 4). We also noticed that the step-and-terrace structures in Fig. 1d form concentric ellipses, and propose that this feature is formed by epitaxial growth on top of thinner single-domain crystal films at a later stage (see Supplementary Fig. 5).

X-ray diffraction data for the films are shown in Fig. 2a and b. The observed out-of-plane diffraction spots are consistent with a molecular layer structure that is parallel to the  $a$  and  $b$  axes. The observation of Bragg reflections up to 14th order indicates that the films have a highly crystalline nature. At high incident angles of the X-rays, we observed 16 diffraction spots that could be ascribed to Bragg reflections with indices that include an in-plane component (Fig. 2b), where the refined unit cell—monoclinic  $P2_1/a$ ,  $a = 5.91(15)$  Å,  $b = 7.88(1)$  Å,  $c = 29.12(19)$  Å,  $\beta = 91.0(8)^\circ$ ,  $V = 1357(4)$  Å<sup>3</sup>—is consistent with that of the bulk crystal<sup>22</sup>. These results provide unambiguous evidence that the films are single-crystalline with a long-range translational symmetry.

The data show that the growth direction is parallel to  $[1\ -1\ 0]$  in many (about 60–70%) of the deposited films (Fig. 1c). On the other

**b**, Micrographs of a  $20 \times 7$  array of inkjet-printed C<sub>8</sub>-BTBT single-crystal thin films. **c**, Crossed Nicols polarized micrographs of the film. **d**, Expanded micrograph of the film, showing stripes caused by molecular-layer steps. **e**, Atomic-force microscopy image and the height profile (below) showing the step-and-terrace structure on the film surfaces.



**Figure 2 | Synchrotron-radiated single-crystal X-ray diffraction and polarized absorption spectra.** Oscillation photographs for out-of-plane diffraction (**a**) and for high-incident-angle diffraction (**b**) of inkjet-printed C<sub>8</sub>-BTBT single-crystal thin films, where  $\omega$  is the incident angle. The Bragg reflections observed in **b** correspond to the indices, which contain in-plane components. The refined unit cell obtained from the reflections is consistent with that of the bulk crystal. **c**, Polarized optical absorption spectra with coefficient  $\alpha$  and with polarization parallel to the  $a$  and  $b$  axes in the single-crystal film, demonstrating optical anisotropy with regard to these principal axes. **d**, View of the molecular arrangement of C<sub>8</sub>-BTBT in the crystal<sup>21</sup>.

hand, the bright-to-dark images observed through the crossed Nicol prisms originate from in-plane optical anisotropy of the single-crystal films. Figure 2c shows the polarized optical absorption spectra of our single-crystal films with the electric field of the light parallel to the *a* or the *b* axis. The spectra show a clear anisotropy in their absorption intensity and peak energy. The absorption intensity is much higher along the *a* axis and peaks at 3.43 eV, whereas the absorption intensity along the *b* axis is comparatively weak and peaks at a higher energy of 3.47 eV. We note that the transition dipole for the lowest electronic excitation between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) is polarized parallel to the molecular plane (Fig. 2c). The difference in the absorption intensity can be clearly ascribed to the orientation of the molecular planes within the *a*–*b* plane. In contrast, the difference in peak energies is due to Davydov splitting along the *a* and *b* axes; this is characteristic of herringbone-type molecular arrangements within single-crystal films, as observed in anthracene<sup>23</sup> or pentacene<sup>24</sup>.

Field-effect devices were fabricated for the single-crystal films with a top-contact/top-gate geometry, composed of 30-nm Au films as the source and drain electrodes, and films of parylene C (capacitance per unit area of  $C = 4.2 \text{ nF cm}^{-2}$ ) as the gate dielectric layers. The typical channel width and length were 145  $\mu\text{m}$  and 100  $\mu\text{m}$ , respectively. The direct-current field-effect characteristics at room temperature (300 K) were measured in an argon-filled glove box. The transfer and output characteristics of this device are shown in Fig. 3. The mobility in the saturation regime reaches  $16.4 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$  on average, and the maximum value is as high as  $31.3 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ . The on/off current ratio is  $10^5$ – $10^7$ , and the subthreshold slope was about 2 V per decade with a threshold voltage of about  $-10 \text{ V}$ . Injection barriers at the source/drain contacts may have remained, as manifested by the slightly nonlinear source/drain current–voltage ( $I_{\text{sd}} - V_{\text{sd}}$ ) dependence at low voltages. Hardly any current hysteresis was observed in the transfer and output characteristics, where the shift in the threshold voltage from forward to reverse sweeps was less than 0.1 V. This

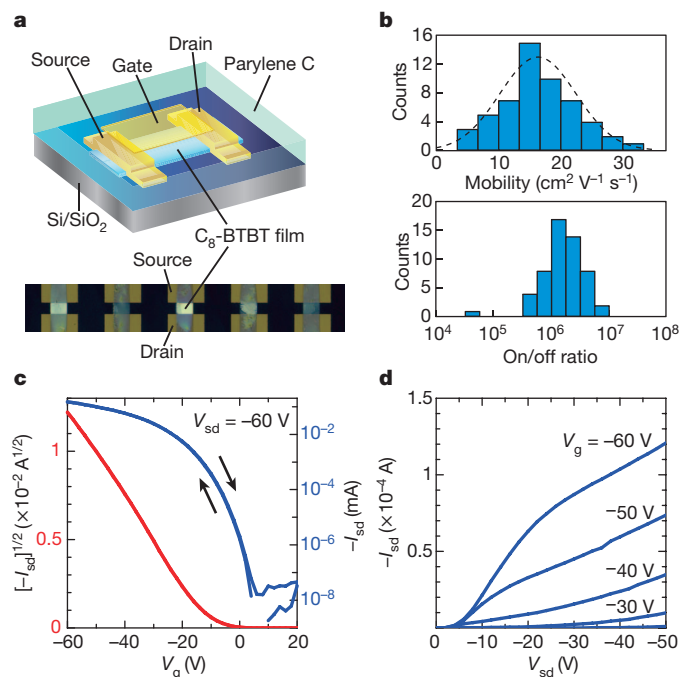
feature is probably associated with the negligible charge-trapping effects between the single-crystal surface of  $\text{C}_8$ -BTBT and the parylene gate dielectric layer. The slope of the transfer curve (Fig. 3c) presents a distinct kink feature, as reported in other organic single-crystal devices<sup>8</sup>, which clearly demonstrates the high quality of the semiconductor–insulator interface. We also found that the characteristics were not influenced by the existence of a few domain boundaries and were not degraded by more than 10% after the films were kept in air for 8 months.

This device performance is much higher than the previous report for  $\text{C}_8$ -BTBT<sup>21</sup> and is comparable to the highest performance obtained (for a rubrene single-crystal device<sup>8</sup>). We consider that the following characteristics of the film formation process are important to achieve high-quality single-crystal film: (1) the liquid–air interfaces need to be ideal locations for diffusion and self-organization of organic molecules (as for Langmuir–Blodgett films<sup>25</sup>) and (2) the gradual growth of single-crystal films is only possible because of the fluidic nature of the microliquid droplet in which laminar flow dominates over turbulent flow<sup>26</sup>. The technique should be applicable to a broad class of functional soluble materials.

The rather broad distribution of apparent mobility (Fig. 3b) indicates that further improvements of our technique should be possible, in areas such as ink composition, the optimization of equipment and the environment, and also subsequent device processing. For example, there is plenty of scope for improving the source/drain contacts. Nonetheless, we believe that this drop-on-demand, non-vacuum and room-temperature printing process of patterned single-crystal semiconductor films is in principle a useful new way of producing transistor arrays on top of plastic substrates, which is indispensable for realizing large-area, light-weight and high-speed electronic products.

Received 20 February; accepted 14 June 2011.

Published online 13 July 2011.



**Figure 3 | Transistor characteristics for the inkjet-printed  $\text{C}_8$ -BTBT single-crystal thin films.** **a**, Schematic of the device structure and micrograph of the thin-film transistors. **b**, Distribution of mobility and on/off ratio measured over 54 transistors. Average mobility is  $16.4 \pm 6.1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ . **c**, Transfer characteristics at  $V_{\text{sd}} = -60 \text{ V}$ . **d**, Output characteristics at various gate voltages  $V_g$ .

- Teal, G. K. Single crystals of germanium and silicon—basic to the transistor and integrated circuit. *IEEE Trans. Electron. Dev.* **23**, 621–639 (1976).
- Tsui, D. C., Stormer, H. L. & Gossard, A. C. Two-dimensional magnetotransport in the extreme quantum limit. *Phys. Rev. Lett.* **48**, 1559–1562 (1982).
- De Poortere, E. P. *et al.* Enhanced electron mobility and high order fractional quantum Hall states in AlAs quantum wells. *Appl. Phys. Lett.* **80**, 1583–1585 (2002).
- Ohtomo, A. & Hwang, H. Y. A high-mobility electron gas at the  $\text{LaAlO}_3/\text{SrTiO}_3$  heterointerface. *Nature* **427**, 423–426 (2004).
- Tsukazaki, A. *et al.* Observation of the fractional quantum Hall effect in an oxide. *Nature Mater.* **9**, 889–893 (2010).
- Sundar, V. C. *et al.* Elastomeric transistor stamps: reversible probing of charge transport in organic crystals. *Science* **303**, 1644–1646 (2004).
- Briseno, A. L. *et al.* Patterning organic single-crystal transistor arrays. *Nature* **444**, 913–917 (2006).
- Takeya, J. *et al.* Very high-mobility organic single-crystal transistors with in-crystal conduction channel. *Appl. Phys. Lett.* **90**, 102120 (2007).
- Yan, H. *et al.* A high-mobility electron-transporting polymer for printed transistors. *Nature* **457**, 679–686 (2009).
- Rivnay, J. *et al.* Large modulation of carrier transport by grain-boundary molecular packing and microstructure in organic thin films. *Nature Mater.* **8**, 952–958 (2009).
- Noh, Y.-Y., Zhao, N., Caironi, M. & Sirringhaus, H. Downscaling of self-aligned, all-printed polymer thin-film transistors. *Nature Nanotechnol.* **2**, 784–789 (2007).
- Whitesides, G. M. & Grzybowski, B. Self-assembly at all scales. *Science* **295**, 2418–2421 (2002).
- Lim, J. A., Lee, H. S., Lee, W. H. & Cho, K. Control of the morphology and structural development of solution-processed functionalized acenes for high-performance organic transistors. *Adv. Funct. Mater.* **19**, 1515–1525 (2009).
- Tung, H.-H., Paul, E. L., Midler, M. & McCauley, J. A. *Crystallization of Organic Compounds: An Industrial Perspective* 179–205 (Wiley, 2009).
- Ebata, H. *et al.* Highly soluble [1]benzothieno[3,2-*b*]benzothiophene (BTBT) derivatives for high-performance, solution-processed organic field-effect transistors. *J. Am. Chem. Soc.* **129**, 15732–15733 (2007).
- Hiraoka, M. *et al.* On-substrate synthesis of molecular conductor films and circuits. *Adv. Mater.* **19**, 3248–3251 (2007).
- Hasegawa, T., Hiraoka, M. & Yamada, T. Double-shot inkjet printing of donor–acceptor-type organic charge-transfer complexes: wet/nonwet definition and its use for contact engineering. *Thin Solid Films* **518**, 3988–3991 (2010).
- Sirringhaus, H. *et al.* High-resolution inkjet printing of all-polymer transistor circuits. *Science* **290**, 2123–2126 (2000).
- Chen, Z., Liu, M., Liu, G.-y. & Tan, L. Evaporation induced two-dimensional buckling within liquid droplet. *Appl. Phys. Lett.* **95**, 223104 (2009).

20. Deegan, R. D. *et al.* Capillary flow as the cause of ring stains from dried liquid drops. *Nature* **389**, 827–829 (1997).
21. Uemura, T., Hirose, Y., Uno, M., Takimiya, K. & Takeya, J. Very high mobility in solution-processed organic thin-film transistors of highly ordered [1]benzothieno[3,2-*b*]benzothiophene derivatives. *Appl. Phys. Expr.* **2**, 111501 (2009).
22. Izawa, T., Miyazaki, E. & Takimiya, K. Molecular ordering of high-performance soluble molecular semiconductors and re-evaluation of their field-effect transistor characteristics. *Adv. Mater.* **20**, 3388–3392 (2008).
23. Pope, M. & Swenberg, C. E. *Electronic Processes in Organic Crystals and Polymers* 2nd edn, 59–66 (Oxford Science, 1999).
24. Faltermeyer, D., Gompf, B., Dressel, M., Tripathi, A. K. & Pflaum, J. Optical properties of pentacene thin films and single crystals. *Phys. Rev. B* **74**, 125416 (2006).
25. Ulman, A. *An Introduction to Ultrathin Organic Films, from Langmuir-Blodgett to Self-Assembly* 1st edn (Academic, 1991).
26. deMello, A. J. Control and detection of chemical reactions in microfluidic systems. *Nature* **442**, 394–402 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to Nippon Kayaku for providing C<sub>8</sub>-BTBT. We thank K. Takimiya and S. Horiuchi for discussions, H. Okamoto and H. Matsuzaki for help with optical measurements, K. Kobayashi for help with the X-ray measurements, and T. Iwade for help with atomic-force microscopy and measurements of device

characteristics. The synchrotron X-ray study was performed with the approval of the Photon Factory Program Advisory Committee (no. 2009S2-003). This work was supported by the New Energy and Industrial Technology Development Organization (NEDO) through a Grant for Industrial Technology Research and also by the Japan Society for the Promotion of Science (JSPS) through its Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program).

**Author Contributions** H. Minemawari was responsible for ink fabrication, inkjet printing, microscopic observations, X-ray diffraction measurements, and measurements of the device characteristics of all the films. T.Y. prepared substrates with the wet/non-wet surface patterning, assisted in inkjet printing and X-ray diffraction measurements, and performed atomic-force microscopy and device characteristics measurements. H. Matsui guided sample preparation and inkjet printing, and conducted DFT molecular orbital calculations. J.T. assisted with X-ray diffraction measurements and performed optical anisotropic absorption measurements. S.H. assisted with optical anisotropic absorption measurements. R.C. assisted in the ink fabrication. R.K. assisted with X-ray diffraction measurements. T.H. conceptualized and directed the research project, and wrote most of the manuscript. All the authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to T.H. (t-hasegawa@aist.go.jp).



# MicroRNA-mediated conversion of human fibroblasts to neurons

Andrew S. Yoo<sup>1†\*</sup>, Alfred X. Sun<sup>2\*</sup>, Li Li<sup>3,4,5\*</sup>, Aleksandr Shcheglovitov<sup>6\*</sup>, Thomas Portmann<sup>6</sup>, Yulong Li<sup>3</sup>, Chris Lee-Messer<sup>7</sup>, Ricardo E. Dolmetsch<sup>6</sup>, Richard W. Tsien<sup>3</sup> & Gerald R. Crabtree<sup>1</sup>

Neurogenic transcription factors and evolutionarily conserved signalling pathways have been found to be instrumental in the formation of neurons<sup>1,2</sup>. However, the instructive role of microRNAs (miRNAs) in neurogenesis remains unexplored. We recently discovered that miR-9\* and miR-124 instruct compositional changes of SWI/SNF-like BAF chromatin-remodelling complexes, a process important for neuronal differentiation and function<sup>3–6</sup>. Nearing mitotic exit of neural progenitors, miR-9\* and miR-124 repress the BAF53a subunit of the neural-progenitor (np)BAF chromatin-remodelling complex. After mitotic exit, BAF53a is replaced by BAF53b, and BAF45a by BAF45b and BAF45c, which are then incorporated into neuron-specific (n)BAF complexes essential for post-mitotic functions<sup>4</sup>. Because miR-9/9\* and miR-124 also control multiple genes regulating neuronal differentiation and function<sup>5,7–13</sup>, we proposed that these miRNAs might contribute to neuronal fates. Here we show that expression of miR-9/9\* and miR-124 (miR-9/9\*-124) in human fibroblasts induces their conversion into neurons, a process facilitated by *NEUROD2*. Further addition of neurogenic transcription factors *ASCL1* and *MYT1L* enhances the rate of conversion and the maturation of the converted neurons, whereas expression of these transcription factors alone without miR-9/9\*-124 was ineffective. These studies indicate that the genetic circuitry involving miR-9/9\*-124 can have an instructive role in neural fate determination.

During the course of exploring the roles of miR-9/9\* and miR-124 (ref. 5), we noted that these miRNAs induced neuronal morphologies in cultured cells. To explore this effect in greater detail, we prepared a single lentiviral vector that expresses both precursors of miR-9/9\* and miR-124 along with a turbo red fluorescent protein (tRFP) marker, and infected human neonatal foreskin fibroblasts (Supplementary Fig. 1). The fibroblast culture was free of neural progenitors, keratinocytes or melanocytes (Supplementary Figs 2–4). Remarkably, fibroblasts expressing miR-9/9\*-124 showed a rapid reduction in proliferation, displayed neuron-like morphologies (Supplementary Fig. 5) and expressed MAP2, a marker of post-mitotic neurons, within 4 weeks after infection (Fig. 1a, left). This was owing to synergism between miR-9/9\* and miR-124, as expressing these miRNAs separately did not lead to the appearance of MAP2-positive cells (Supplementary Fig. 6). In light of the low percentage of MAP2-positive cells obtained with miRNAs only (less than 5%, Fig. 1b), we began adding neurogenic transcription factors and found that *NEUROD2* (refs 14–17) was most effective at increasing the conversion frequency (Supplementary Fig. 7). We estimate that ~50% of these cells have acquired neuronal fates as indicated by MAP2 expression 30 days after infection (Fig. 1a, right, b). However, because cells detached, remained uninfected or died during the conversion process, a conservative estimate is that ~5% of the starting cells became neurons. Importantly, neither *NEUROD2* alone nor non-specific miRNA (miR-NS) could convert fibroblasts into neurons

(Fig. 1b), demonstrating the essential role of miR-9/9\*-124 in this process. Synergism between miR-9/9\* and miR-124 seemed to be crucial: expressing miR-9/9\* and miR-124 individually with *NEUROD2* failed to produce MAP2-positive cells (Supplementary Fig. 6). Using EdU-incorporation, we found that miR-9/9\*-124-infected fibroblasts had exited the cell cycle 1 week after infection (Supplementary Fig. 8), consistent with the anti-proliferative role of these miRNAs<sup>5</sup>. Lastly, immunostains indicated that the induced neurons expressed *SCN1a*, a key contributor to neuronal excitability, as well as synapsin 1 and NMDA receptor 1 (Fig. 1c).

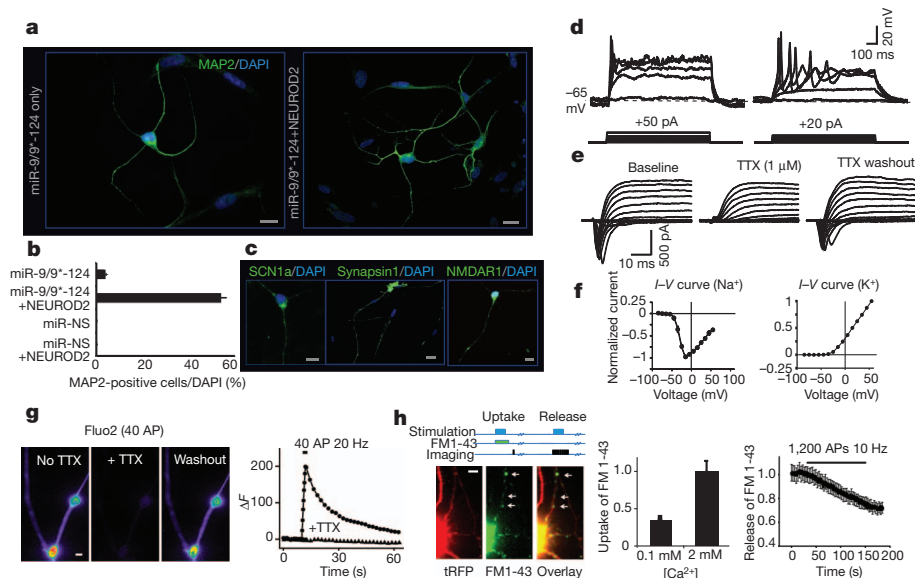
Using whole-cell patch recording, we found that injecting depolarizing current in induced neurons (cultured up to 8 weeks) could consistently trigger single action potentials and in some cases, repetitive firing (Fig. 1d). Moreover, their resting membrane potential ( $-34.1 \pm 1.7$  mV; Supplementary Fig. 9) was significantly more negative than that of control fibroblasts ( $-20.4 \pm 0.6$  mV,  $n = 4$ ). Applying a series of voltage steps to the induced cells evoked large inward currents closely followed by outward currents, which were not observed in the fibroblasts (Supplementary Fig. 10). Importantly, adding 1  $\mu$ M tetrodotoxin (TTX) completely and reversibly blocked the initial inward current, confirming that the current was due to voltage-gated sodium channels (Fig. 1e), as would be expected from the current–voltage ( $I$ – $V$ ) curve of inward currents (Fig. 1f, left). The  $I$ – $V$  curve of outward currents showed the characteristics of voltage-gated potassium channels in neurons (Fig. 1f, right). Moreover, some of these cells exhibited postsynaptic currents, which could be reversibly blocked by 2,3-dihydroxy-6-nitro-7-sulfamoyl-benzo[*f*]quinoxaline-2,3-dione (NBQX) and 2-amino-5-phosphonopentanoic acid (APV) (Supplementary Fig. 11).

We examined the ability of cells converted by miR-9/9\*-124-*NEUROD2* to elicit a stimulation-dependent calcium influx using the calcium indicator Fluo2. Field stimulation triggered calcium influx that could be abolished by adding TTX (Fig. 1g) or 200  $\mu$ M  $\text{Cd}^{2+}$  (Supplementary Fig. 12), demonstrating the ability of converted cells to support activity-dependent  $\text{Ca}^{2+}$  influx through voltage-gated  $\text{Ca}^{2+}$  channels without any requirement for a pre-pulse. Activity-dependent uptake and release of the lipophilic dye FM1-43 was used to evaluate the ability to form functional presynaptic terminals<sup>18</sup>. We found that the induced cells were able to take up and release FM dyes in a stimulation-dependent and  $\text{Ca}^{2+}$ -dependent manner (Fig. 1h).

Because the miR-9/9\*-124-*NEUROD2*-induced cells only occasionally showed repetitive action potentials, we sought to optimize the maturation of the cells by introducing additional neurogenic factors. Because *ASCL1* and *MYT1L* were previously shown to be important for converting mouse embryonic fibroblasts into functionally mature neurons<sup>19</sup>, we expressed miR-9/9\*-124 together with *NEUROD2*, *ASCL1* and *MYT1L* (DAM). We found that the miR-9/9\*-124-DAM-converted cells were positive for MAP2 expression in approximately

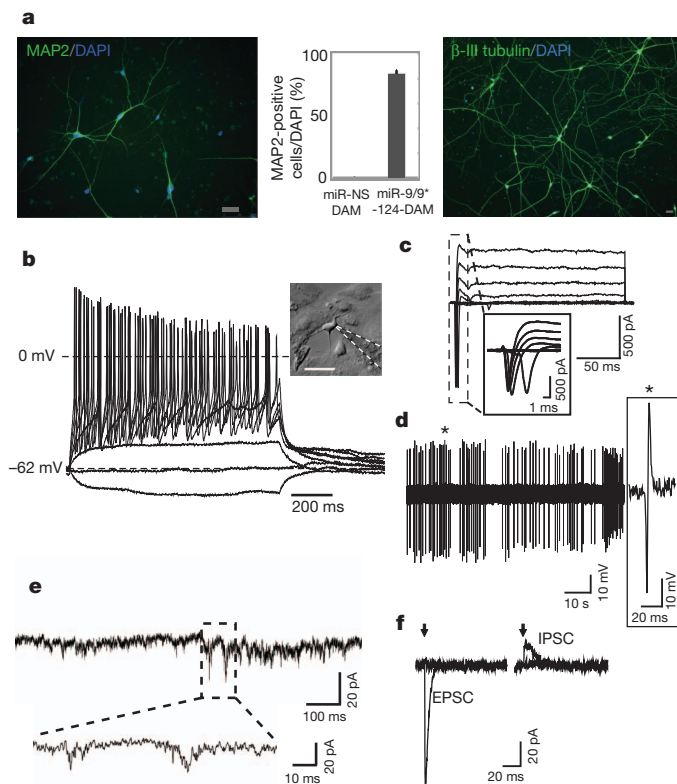
<sup>1</sup>Howard Hughes Medical Institute and the Departments of Developmental Biology and of Pathology, Stanford University, Stanford, California 94305, USA. <sup>2</sup>Program in Cancer Biology, Stanford University, Stanford, California 94305, USA. <sup>3</sup>Department of Molecular and Cellular Physiology, Stanford University, Stanford, California 94305, USA. <sup>4</sup>Medical Scientist Training Program, Stanford University, Stanford, California 94305, USA. <sup>5</sup>Neuroscience Program, Stanford University, Stanford, California 94305, USA. <sup>6</sup>Department of Neurobiology, Stanford University, Stanford, California 94305, USA. <sup>7</sup>Department of Neurology, Stanford University, Stanford, California 94305, USA. <sup>†</sup>Present address: Department of Developmental Biology, Washington University in St Louis, St Louis, Missouri 63110, USA.

\*These authors contributed equally to this work.



**Figure 1 | miRNA-induced transformation of human fibroblasts.** **a**, MAP2 expression in miR-9/9\*-124 only (left) and miR-9/9\*-124-NEUROD2 (right) converted cells. DAPI, 4',6-diamidino-2-phenylindole. Scale bar, 20  $\mu$ m. **b**, Quantification of MAP2-positive cells with processes at least three times the length of the cell body from ten random fields. The graph represents the percentage of MAP2-positive cells over DAPI-positive cells. miR-9/9\*-124 only:  $n = 558$ ; miR-9/9\*-124-NEUROD2:  $n = 658$  cells. The error bars are s.e.m. MAP2 signal was undetectable in fibroblasts infected with miR-NS, with or without NEUROD2. Scale bar, 20  $\mu$ m. **c**, Expression of SCN1a, synapsin 1 and NMDAR1 in miR-9/9\*-124-NEUROD2-converted cells. Scale bar, 20  $\mu$ m. **d**, Representative traces of action potentials recorded in current clamp in miR-9/9\*-124-NEUROD2-converted cells. Twelve out of 22 cells showed single action potentials and 2 cells showed repetitive firing. **e**, A representative example of a series of voltage steps applied to an miR-9/9\*-124-NEUROD2-induced neuron held at  $-70$  mV. An inward current was observed, blocked by  $1 \mu$ M TTX, and reversed after TTX washout ( $n = 6$ ). **f**,  $I$ - $V$  curve for the peak

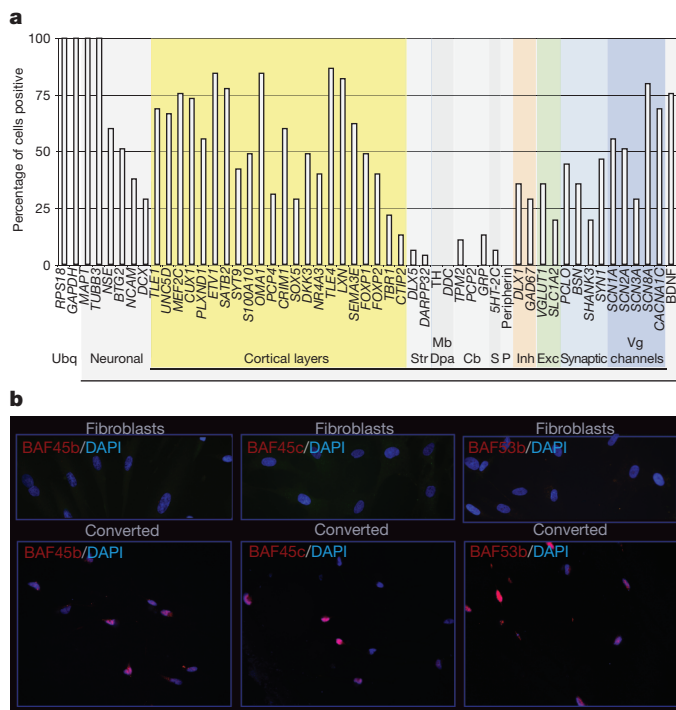
inward (left) and outward (right) currents. **g**, An example of  $\text{Ca}^{2+}$  influx in induced neurons as measured by Fluo2-AM imaging. Images show the peak Fluo2-AM signal on stimulation before and during TTX application, and after TTX washout, respectively. The graph represents the changes in Fluo2 signal over time (circles, no TTX; triangles, with TTX). AP, action potential. Field stimulation is indicated by the black bar. Scale bar, 2  $\mu$ m. **h**, An example of vesicle recycling measured by FM1-43 imaging in induced neurons. Top diagram illustrates the protocol of FM uptake or release experiments. The images represent the typical FM1-43 dye uptake signal (middle) in a converted cell marked by tRFP (left). The left graph shows the measurement of FM1-43 loading in  $2 \text{ mM } \text{Ca}^{2+}$ , which was significantly reduced in low  $\text{Ca}^{2+}$  concentration ( $0.1 \text{ mM}$ ). The right graph quantifies FM1-43 release (de-staining) during stimulations. FM1-43 signal was measured from  $n > 600$  boutons (arrows indicate examples of boutons) from 4 cultures. All error bars are s.e.m. In some cases, the s.e.m. is too small to be resolved. Scale bar, 4  $\mu$ m.



**Figure 2 | Additional neural factors enhance the conversion to neurons.** **a**, MAP2 (left) and  $\beta$ -III tubulin (right) immunostaining of miR-9/9\*-124-DAM-converted cells. Scale bar, 40  $\mu$ m. The graph represents the percentage of MAP2-positive cells over DAPI-positive cells. The error bars are s.e.m.;  $n = 150$  cells. **b**, A representative current clamp recording from a cell with typical neuronal morphology (see inset; scale bar, 50  $\mu$ m). Voltage deflections were elicited by somatic current injections of various amplitudes ( $\Delta = 5$  pA). **c**, A representative voltage clamp recording of the net current at various membrane potentials ( $-40$  to  $+20$  mV,  $\Delta V = 10$  mV,  $V_{\text{hold}} = -90$  mV). **d**, A representative trace of spontaneously active cells recorded in cell-attached mode. **e**, A representative trace demonstrating spontaneous EPSCs. **f**, Representative traces of evoked postsynaptic currents (left, EPSC; right, IPSC) obtained in response to local field stimulation with single current pulses (1 ms) of various amplitudes (left, 0.25 and 0.3 mA; right, 0.3 and 0.4 mA) at different membrane holding potentials (left,  $-70$  mV; right, 0 mV). The arrows indicate the time when stimulation was applied. Stimulation artefacts were eliminated for clarity.

80% of the cells remaining on the coverslips (Fig. 2a, representing ~10% of the initially plated cells), and showed extensive neurite outgrowth as illustrated by  $\beta$ -III tubulin staining (Fig. 2a, right). In addition, miR-9/9\*-124-DAM resulted in complete exit from cell cycle as assayed by EdU pulsing for 4 days (0/176 positive) whereas nearly all control cells were positive for EdU (97/107). Importantly, DAM factors with miRNA-NS failed to produce neurons as assayed by MAP2 staining (Supplementary Fig. 13). About 80% of the miR-9/9\*-124-DAM-converted cells were able to fire repetitive action potentials in response to current injections, and showed typical sodium and potassium currents during voltage clamp depolarizations (Fig. 2b–c and Supplementary Figs 9 and 14). Among recorded cells, we also observed spontaneously active cells (2/21) (Fig. 2d). Spontaneous excitatory postsynaptic currents (EPSCs) were seen in 10 out of 14 induced cells (Fig. 2e) without co-cultured primary neurons (Supplementary Figs 9 and 15a). Furthermore, the induced cells exhibited evoked EPSCs and inhibitory postsynaptic currents (IPSCs) in response to local stimulation (Fig. 2f). Importantly, neuronal identity was stable after the removal of exogenous expression of miR-9/9\*-124 and DAM after 3 weeks of induction, as they still stained positive for SV2 and synapsin 1 (Supplementary Fig. 16).

We next performed single-cell analysis to characterize the types of neurons in miR-9/9\*-124-DAM-induced cells. From randomly collected single cells 4 weeks post-infection, we analysed a total of 45 induced neurons (based on *MAPT* and *TUBB3* co-expression) for genes expressed in different types of neurons. We found that most induced cells were positive for genes expressed in cortical layers (Fig. 3a and Supplementary Fig. 17). Interestingly, we did not detect a peripheral nervous system marker (peripherin) or dopaminergic/noradrenergic markers (*DDC*, *TH*). Striatal markers (*DLX5* and *DARPP32* (also known as *PPP1R1B*)), the serotonergic marker *5HT-2C* (also known as *HTR2C*), and cerebellar genes (*PCP2*, *GRP*, *TPM2*)

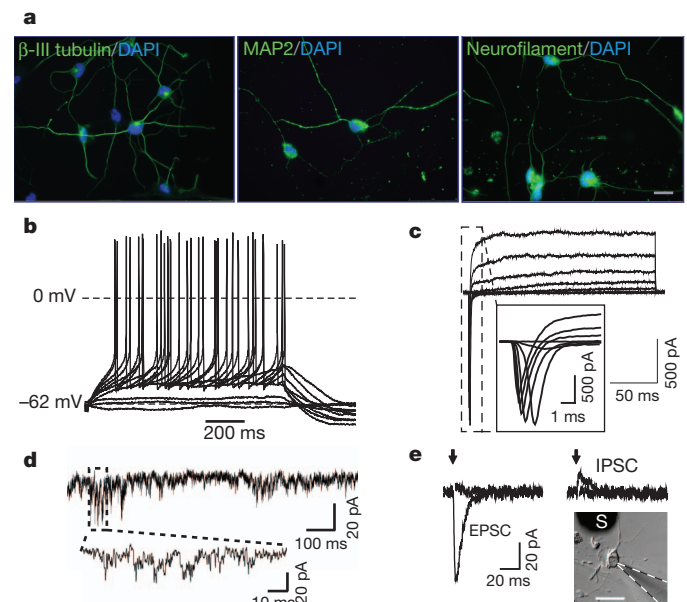


**Figure 3 | Characterization of induced neurons and nBAF subunit expression.** **a**, Multiplex quantitative polymerase chain reaction (qPCR) of 45 induced neurons for genes of specific brain regions and cell types. Cb, cerebellum; Dpa, dopaminergic; Exc, excitatory; Inh, inhibitory; Mb, midbrain; P, peripheral nervous system; S, serotonergic; Str, striatum; Ubq, ubiquitous; Vg, voltage gated. **b**, Fibroblasts stained negative for neuron-specific subunits of BAF complexes (top) whereas induced neurons expressed BAF45b, BAF45c and BAF53b (bottom). Scale bar, 20  $\mu$ m.

were expressed in only a small number of cells. The miR-9/9\*-124-DAM-induced cells seemed to be heterogeneously excitatory (*VGLUT1* (also known as *SLC17A7*) and *SLC1A2*) and inhibitory (*GAD67* (also known as *GAD1*) and *DLX1*) (Fig. 3a), which is further supported by immunostains of *VGLUT1* and *GABA* (Supplementary Fig. 18). Moreover, by 4 weeks post-infection, the induced cells already expressed genes important for synaptic structure and function, including *SYN1*, *BSN*, *PCLO* and *SHANK3* (Fig. 3a).

miR-9\* and miR-124 target separate sites in the 3' untranslated region (UTR) of BAF53a, a subunit of BAF complexes resulting in repression of BAF53a and activation of BAF53b, which is involved in an evolutionarily conserved program of neural development<sup>20</sup>. Remarkably, we found that all of the nBAF subunits (BAF53b, 45b and 45c) were induced in the converted cells (Fig. 3b). In embryonic stem cells, BAF complexes function across the genome at several thousand sites to control placement of polycomb repressive complex 2 and the H3K27me3 repressive mark<sup>21,22</sup>. Hence, one role of the miRNAs might be to induce stable epigenetic changes involving polycomb function across the genome.

In addition to BAF53a, miR-9/9\* and miR-124 also target other genes essential for neurogenesis and neuronal functions<sup>23</sup> including components of the REST complex such as REST and CoREST<sup>8,9,24–26</sup>, and PTBP-1 (ref. 7). We found that human fibroblasts expressed BAF53a, which could be repressed by miR-9/9\*-124 (Supplementary Fig. 19). However, prolonging the expression of BAF53a only incompletely blocked neuronal conversion of fibroblasts, as assayed by MAP2 staining (data not shown). Prolonging the expression of REST, CoREST or PTBP1 yielded similar results (data not shown). These findings indicate that in inducing cell fate transformations, the miRNAs miR-9/9\* and miR-124 operate programmatically on multiple targets.



**Figure 4 | Conversion of adult fibroblasts by miR-9/9\*-124-DAM.**

**a**, Immunostaining of  $\beta$ -III tubulin (left), MAP2 (middle) and neurofilament (right) in dermal fibroblasts of a 30-year-old individual converted by miR-9/9\*-124-DAM. Scale bar, 20  $\mu$ m. **b**, A representative current clamp recording. Voltage deflections were elicited by somatic current injections of various amplitudes ( $\Delta = 2$  pA). **c**, A representative voltage clamp recording of the net current at various membrane potentials ( $-40$  to  $+20$  mV,  $\Delta V = 10$  mV,  $V_{\text{hold}} = -90$  mV). **d**, A representative trace of spontaneous EPSCs. **e**, Representative traces of evoked postsynaptic currents obtained from converted adult cells (see inset; scale bar, 50  $\mu$ m) in response to local field stimulation with single current pulses (1 ms) of various amplitudes (left, 0.2 and 0.25 mA; right, 0.4 and 0.5 mA). Evoked EPSPs and IPSCs were recorded at  $-70$  mV and  $+30$  mV in the presence of picrotoxin and NBQX/APV, respectively. S, stimulator.



Lastly, we asked whether our approach could be effective in converting adult fibroblasts. We found that adult human dermal fibroblasts (from a 30-year-old female) could be converted into neurons (Fig. 4a), albeit more slowly. Recordings from adult cells converted by miR-9/9\*-124-DAM 6 weeks after infection showed that they were able to generate action potentials (Fig. 4b). They also demonstrated typical voltage-gated sodium and potassium currents (Fig. 4c), spontaneous EPSCs (Fig. 4d and Supplementary Fig. 15b) and evoked EPSCs and IPSCs (Fig. 4e) without co-cultured primary neurons.

Our studies show that activating a neural developmental regulatory circuit involving miRNAs in human fibroblasts can surprisingly induce their conversion into neurons, indicating an instructive role for this circuitry in human neurogenesis. In our study, neurogenic transcription factors, delivered either singly (*NEUROD2*), or in combination (*NEUROD2*, *ASCL1* and *MTYL1*), seem to function synergistically with the neurogenic activities of miR-9/9\*-124. This raises the possibility of inducing various types of neurons using miR-9/9\*-124 together with different sets of transcriptional factors.

## METHODS SUMMARY

**Transduction of human fibroblasts.** A synthetic cluster of miR-9/9\* and miR-124 validated previously to express miR-9\* and miR-124 (ref. 5) was inserted downstream of tRFP in the pLentiMIR lentiviral construct carrying a puromycin selection cassette (Open Biosystems) driven by either a CMV promoter or a doxycycline-responsive promoter. A non-silencing sequence, which produces miRNA-NS, was used as a control (Open Biosystems). Each transcription factor was cloned downstream of the EF1 $\alpha$  promoter in a separate lentiviral construct. Typically, infected human fibroblasts were maintained in fibroblast media for 3–4 days before selection with appropriate antibiotics in Neuronal Media (ScienCell) supplemented with VPA (1 mM) and basic FGF (20 ng ml<sup>-1</sup>). dbcAMP (500  $\mu$ M) was added 10 days later to enhance cell survival. Human BDNF and NT3 (10 ng ml<sup>-1</sup>; Peprotech) were added to the media after 3–4 weeks. Media were changed every 4 days.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 21 July 2010; accepted 27 June 2011.

Published online 13 July 2011.

- Hansen, D. V., Lui, J. H., Parker, P. R. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).
- Hansen, D. V., Rubenstein, J. L. & Kriegstein, A. R. Deriving excitatory neurons of the neocortex from pluripotent stem cells. *Neuron* **70**, 645–660 (2011).
- Lessard, J. et al. An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* **55**, 201–215 (2007).
- Wu, J. et al. Regulation of dendritic development by neuron-specific chromatin remodeling complexes. *Neuron* **56**, 94–108 (2007).
- Yoo, A. S., Staahl, B. T., Chen, L. & Crabtree, G. R. MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. *Nature* **460**, 642–646 (2009).
- Wu, J. I., Lessard, J. & Crabtree, G. R. Understanding the words of chromatin regulation. *Cell* **136**, 200–206 (2009).
- Makeyev, E. V., Zhang, J., Carrasco, M. A. & Maniatis, T. The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell* **27**, 435–448 (2007).
- Packer, A. N., Xing, Y., Harper, S. Q., Jones, L. & Davidson, B. L. The bifunctional microRNA miR-9/miR-9\* regulates REST and CoREST and is downregulated in Huntington's disease. *J. Neurosci.* **28**, 14341–14346 (2008).
- Visvanathan, J., Lee, S., Lee, B., Lee, J. W. & Lee, S. K. The microRNA miR-124 antagonizes the anti-neural REST/SCP1 pathway during embryonic CNS development. *Genes Dev.* **21**, 744–749 (2007).
- Cheng, L. C., Pastrana, E., Tavazoie, M. & Doetsch, F. miR-124 regulates adult neurogenesis in the subventricular zone stem cell niche. *Nature Neurosci.* **12**, 399–408 (2009).

- Krichevsky, A. M., Sonntag, K. C., Isacson, O. & Kosik, K. S. Specific microRNAs modulate embryonic stem cell-derived neurogenesis. *Stem Cells* **24**, 857–864 (2006).
- Maiorano, N. A. & Mallamaci, A. Promotion of embryonic cortico-cerebral neurogenesis by miR-124. *Neural Develop.* **4**, 40 (2009).
- Tang, X. et al. A simple array platform for microRNA analysis and its application in mouse tissues. *RNA* **13**, 1803–1822 (2007).
- Lin, C. H. et al. The dosage of the neuroD2 transcription factor regulates amygdala development and emotional learning. *Proc. Natl Acad. Sci. USA* **102**, 14877–14882 (2005).
- McCormick, M. B. et al. *neuroD2* and *neuroD3*: distinct expression patterns and transcriptional activation potentials within the *neuroD* gene family. *Mol. Cell. Biol.* **16**, 5792–5800 (1996).
- Olson, J. M. et al. *NeuroD2* is necessary for development and survival of central nervous system neurons. *Dev. Biol.* **234**, 174–187 (2001).
- Ince-Dunn, G. et al. Regulation of thalamocortical patterning and synaptic maturation by *NeuroD2*. *Neuron* **49**, 683–695 (2006).
- Ryan, T. A. et al. The kinetics of synaptic vesicle recycling measured at single presynaptic boutons. *Neuron* **11**, 713–724 (1993).
- Vierbuchen, T. et al. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
- Parrish, J. Z., Kim, M. D., Jan, L. Y. & Jan, Y. N. Genome-wide analyses identify transcription factors required for proper morphogenesis of *Drosophila* sensory neuron dendrites. *Genes Dev.* **20**, 820–835 (2006).
- Ho, L. et al. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc. Natl Acad. Sci. USA* **106**, 5187–5191 (2009).
- Ho, L., Miller, E. L., Ronan, J. L., Ho, W. Q., Jothi, R. & Crabtree, G. R. esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signaling and by regulating polycomb function. *Nature Cell Biol.* (in the press).
- Coolen, M. & Bally-Cuif, L. MicroRNAs in brain development and physiology. *Curr. Opin. Neurobiol.* **19**, 461–470 (2009).
- Wu, J. & Xie, X. Comparative sequence analysis reveals an intricate network among REST, CREB and miRNA in mediating neuronal gene expression. *Genome Biol.* **7**, R85 (2006).
- Laneve, P. et al. A minicircuitry involving REST and CREB controls miR-9-2 expression during human neuronal differentiation. *Nucleic Acids Res.* **38**, 6895–6905 (2010).
- Andres, M. E. et al. CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proc. Natl Acad. Sci. USA* **96**, 9873–9878 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank I. Graef and A. Cho for helpful suggestions and reagents, A. Kuo and W. Ho for technical help, and X. Bao and P. Khavari for their generous gift of reagents. A.S.Y. is a fellow of the Helen Hay Whitney Foundation. A.X.S. is funded by the Agency of Science, Technology and Research of Singapore (A\*STAR). L.L. is supported by the Stanford Medical Scientist Training Program, National Institutes of Mental Health (NIMH) F30MH093125, and the Frances B. Nelson predoctoral fellowship. A.S. is supported by the CIRM post-doctoral fellowship. T.P. is supported by a Swiss National Science Foundation SNSF fellowship for advanced researchers (PA00P3\_134196). R.E.D. is supported by the NIH Director's Award, and awards from the Simon's Foundation and the CIRM. R.E.D. is also grateful for funding from B. and F. Horowitz, M. McCaffrey, B. and J. Packard, P. Kwan and K. Wang. R.W.T. is supported by grants from the Simons, Mathers and Burnett Family Foundations. This work was supported by grants from the Howard Hughes Medical Institute (G.R.C.) and the NIH (HD55391, AI060037 and NS046789 to G.R.C., and NS24067, GM58234 and MH064070 to R.W.T.).

**Author Contributions** A.S.Y., A.X.S., and G.R.C. generated the hypotheses and designed experiments. A.S.Y. and A.X.S. performed experiments, generated data in all figures and Supplementary Data. A.S. and L.L. designed and performed experiments for Figs 1, 2 and 4 and Supplementary Data. T.P. designed and performed experiments in Fig. 3a. Y.L. generated data presented in Fig. 1. C.L.-M. performed experiments for Supplementary Data. A.S.Y., A.X.S., L.L., A.S., Y.L., T.P., R.W.T., R.E.D. and G.R.C. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to G.R.C. ([crabtree@stanford.edu](mailto:crabtree@stanford.edu)) or A.S.Y. ([yooa@wustl.edu](mailto:yooa@wustl.edu)).

## METHODS

**Plasmid construction and viral preparation.** We have previously constructed a synthetic cluster that expresses the precursors of both miR-9/9\* (NCBI and miRBASE accession numbers MIMAT0000441 and MIMAT0000442) and miR-124 (accession number MIMAT0000422) and validated its ability to generate mature miRNAs of both. Here we cloned this cluster downstream of a tRFP marker into pLemiR (Open Biosystems), driven by either a CMV promoter or a doxycycline-responsive promoter. A non-silencing sequence in pLemiR (miR-NS, which produces non-specific miRNA) was used as a control (Open Biosystems). cDNA of each neural transcription factor used in this study was cloned downstream of the EF1 $\alpha$  promoter in a separate lentiviral construct with blasticidin or neomycin selection. For doxycycline experiments, human fibroblasts were first infected with a lentiviral construct expressing rtTA under the EF1 $\alpha$  promoter and stably selected with hygromycin. Infectious lentiviruses were collected 36–60 h after transfection of Lenti-X 293T cells (Clontech) with appropriate amounts of lentiviral vectors, psPAX2 and pMD2.G (Addgene) using Eugene HD (Roche).

**Cell culture.** All fibroblast cultures (human neonatal foreskin fibroblasts (ATCC, PCS-201-010) and adult dermal fibroblast (ScienCell)) were maintained in fibroblast media (Dulbecco's Modified Eagle Medium; Invitrogen) containing 10% fetal bovine serum (FBS; Omega Scientific),  $\beta$ -mercaptoethanol (Sigma-Aldrich), non-essential amino acids, sodium pyruvate, GlutaMAX, and penicillin/streptomycin (all from Invitrogen). The day before lentiviral infection, human fibroblasts were seeded onto gelatin-coated 24-well tissue culture dishes (MidSci). Next day, cells were infected with filtered viral supernatants in the presence of polybrene ( $8 \mu\text{g ml}^{-1}$ ) overnight. Fresh media were then replaced for 2–3 days with appropriate antibiotics to select for infected cells. Four days after infection, the media was changed to Neuronal Media (ScienCell) supplemented with VPA (1 mM), basic FGF ( $20 \text{ ng ml}^{-1}$ ) with media changes every 4 days. Our experiments to optimize the conversion protocol indicated that VPA and bFGF are beneficial for conversion for the first 2–3 weeks. We also added dbcAMP (0.5 mM, Sigma) after 2 weeks to the media as we found it enhanced cell survival. Human BDNF and NT3 ( $10 \text{ ng ml}^{-1}$ , Peprotech) were added to the media after 3–4 weeks to promote the survival of the induced cells. To facilitate immunostaining and electrophysiological studies in some experiments, cells were trypsinized (0.05% trypsin, Invitrogen) at about 10 days after infection and re-plated onto poly-ornithine (Sigma-Aldrich)/laminin (Roche)/fibronectin (Sigma-Aldrich)-coated glass coverslips.

**Immunofluorescence.** The following antibodies were used for the immunofluorescence studies: mouse anti-MAP2 (Sigma-Aldrich, 1:750), chicken anti-MAP2 (Abcam, 1:30,000), mouse anti- $\beta$ -III tubulin (Covance, 1:30,000), rabbit anti-VGLUT1 (Synaptic Systems, 1:2,000), rabbit anti-SCN1a (Abcam, 1:1,000), rabbit anti-NMDAR1 (Abcam, 1:2,000), rabbit anti-neurofilament 200 (Sigma-Aldrich, 1:2,000), mouse anti-SV2 (Developmental Studies Hybridoma Bank, 1:100), rabbit anti-GABA (Sigma, 1:2,000) and rabbit anti-synapsin1 (Cell Signaling, 1:200). Antibodies against BAF subunits were generated in our laboratory and used as the following concentrations: BAF45b (1:250), BAF45c (1:1,000) and BAF53b (1:500). The secondary antibodies were goat anti-rabbit or mouse IgG conjugated with Alexa-488, -568 or -647 (Invitrogen). For SCN1a and BAF53b staining, biotinylated secondary antibodies were detected using TSA amplification kit (Invitrogen). EdU incorporation assay was performed according to the manufacturer's protocols (Invitrogen). Images were captured using a Leica DM5000B microscope with Leica Application Suite (LAS) Advanced Fluorescence 1.8.0 and a Leica DMI4000B microscope with LAS v.2.8.1.

**Electrophysiology.** Recordings were performed on fibroblasts 5–8 weeks after infection for both miR-9/9\*-124-NEUROD2 and miR-9/9\*-124-DAM converted cells, which were co-cultured with mouse glia. Data were acquired in whole-cell mode at room temperature ( $25^\circ\text{C}$ ) using an Axopatch 200B amplifier (Molecular Devices) or EPC 10 amplifier (HEKA) and sampled at 5 kHz with a 2 kHz low-pass filter. Recording pipette resistance was 2–6 M $\Omega$ . Intrinsic neuronal properties were studied using the following solutions (in mM): extracellular, 140 NaCl, 2.5 KCl, 2.5 CaCl<sub>2</sub>, 2 MgCl<sub>2</sub>, 1 NaH<sub>2</sub>PO<sub>4</sub>, 20 glucose, 10 HEPES, pH 7.4; intracellular, 120 KGluc, 20 KCl, 4 NaCl, 4 Mg<sub>2</sub>ATP, 0.3 NaGTP, 10 Na<sub>2</sub>PCr, 0.5 EGTA, 10 HEPES, pH 7.25. Synaptic activity was measured using the same extracellular solution, supplemented with 50  $\mu\text{M}$  APV, and the following intracellular solution (in mM):

135 CsMeS, 5 CsCl, 10 HEPES, 0.5 EGTA, 1 MgCl<sub>2</sub>, 4 Mg<sub>2</sub>ATP, 0.4 NaGTP, 5 QX-314, pH 7.4 CsOH. Correspondingly, EPSCs were measured at  $-70 \text{ mV}$  whereas IPSCs were measured at either 0 or  $+30 \text{ mV}$  (with 10  $\mu\text{M}$  NBQX and 50  $\mu\text{M}$  APV,  $E_{\text{Cl}^-}$  approximately  $-79 \text{ mV}$ ). Evoked postsynaptic currents were elicited by a stimulating electrode (CBAEC75, FHC) positioned 100–150  $\mu\text{m}$  aside from the cell soma through which brief (1 ms) unipolar current pulses of various amplitudes (0.1–0.9 mA,  $\Delta = 0.05$ –0.1 mA) were applied. Recordings were filtered at 2 kHz and digitally sampled at 10 kHz. Data were collected and initially analysed with Clampfit 10 or the Patchmaster software (HEKA). Further analysis was performed using IgorPRO and MS Excel. Series resistance was left uncompensated owing to the fragility of the cells, but was corrected in the current clamp calculations. The liquid junction potential was calculated to be 15 mV (Clampfit) and corrected in calculating resting membrane potentials according to published methods<sup>27</sup>.

**Calcium imaging.** Cells were loaded with Fluo2-AM (5  $\mu\text{M}$ , TEFLABS) in Tyrode solution (150 mM NaCl, 4 mM KCl, 2 mM CaCl<sub>2</sub>, 2 mM MgCl<sub>2</sub>, 10 mM glucose, 10 mM HEPES, 310–315 mOsm, with pH at 7.35) for 30 min in a  $37^\circ\text{C}$  incubator. After two washes with Tyrode, cells were imaged using a filter cube (excitation  $470 \pm 20 \text{ nm}$  and emission  $535 \pm 50 \text{ nm}$ ). In some cases, 1  $\mu\text{M}$  TTX or 200  $\mu\text{M}$  CdCl<sub>2</sub> was superfused. All images were converted to TIFF files and analysed off-line with Metamorph or ImageJ. All error bars represent s.e.m. For analysis of FM-positive puncta, 1.3  $\mu\text{m}$  in diameter regions of interest were used to cover functional boutons. Photobleaching was corrected by fitting the pre-stimulation baseline to a linear curve.

**FM1-43 imaging.** Cells were superfused with Tyrode solution. Switching of superfusion solution was carried out with a precision of  $<2 \text{ s}$ . Solutions contained 10  $\mu\text{M}$  NBQX and 50  $\mu\text{M}$  D-APV (Tocris Bioscience) to prevent possible recurrent activity and synaptic plasticity. All experiments were performed at room temperature and neurons were stimulated with platinum electrodes. Putative presynaptic boutons were stained with 8  $\mu\text{M}$  FM1-43 (Molecular Probes) using field stimulation for 120 s at 10 Hz, followed by 60 s without stimulation to maximize the loading. In some experiments, 0.1 mM CaCl<sub>2</sub> was used to test the calcium dependency. After 10 min of washing with dye-free Tyrode's solution, individual boutons were destained by field stimulation. FM1-43 dyes were excited at 470 nm (D470-40x; Chroma) and their emission was collected at 535 nm (535/50m). tRFP was excited at 535 nm (535/50ex) and its emission was collected at 580 nm (580 lp). All images were taken at a frame rate of 1–3 Hz by a Cascade 512B camera.

**Single-cell qPCR.** Single cells were collected by clone FACS sorting using a BD influx sorter (BD Biosciences) in 10  $\mu\text{l}$  of a pre-amplification mix containing 40 nM of all primers for genes of interest, and the following components of the CellsDirect One-Step qRT-PCR Kit (Invitrogen): 2 $\times$  Reaction Mix, SuperScript III RT/Platinum Taq Mix. After sorting, samples were reverse transcribed and pre-amplified for 18 cycles. Pre-amplified samples were diluted (3 $\times$ ) with TE buffer and stored at  $-20^\circ\text{C}$ . Sample and assay (primer pairs) preparation for 96.96 Fluidigm Dynamic arrays was done according to the manufacturer's recommendation. Briefly, sample was mixed with 20 $\times$  DNA binding dye sample loading reagent (Fluidigm), 20 $\times$  EvaGreen (Biotium) and TaqMan Gene Expression Master Mix (Applied Biosystems). Assays were mixed with 2 $\times$  assay loading reagent (Fluidigm) and TE to a final concentration of 5  $\mu\text{M}$ . The 96.96 Fluidigm Dynamic Arrays (Fluidigm) were primed and loaded on an IFC Controller HX (Fluidigm) and qPCR experiments were run on a Biomark System for Genetic Analysis (Fluidigm). Data were collected and analysed using the Fluidigm Real-Time PCR Analysis software (v.2.1.3 and v.3.0.2). Melting curves were used to determine specificity of each reaction. Further data analysis was performed using Microsoft Excel. In addition to collected single-cell material, every experiment contained samples for four standard dilutions of a mixed human cDNA library. The collected cells were confirmed based on *RSG18* (18S small ribosomal subunit) and *GAPDH* co-expression. Of these, induced neurons were identified by co-expression of two general neuronal genes *MAPT* and *TUBB3* for further analysis of genes specific to brain regions and cell types.

27. Barry, P. H. JPCalc, a software package for calculating liquid junction potential corrections in patch-clamp, intracellular, epithelial and bilayer measurements and for correcting junction potential measurements. *J. Neurosci. Methods* **51**, 107–116 (1994).